MATHEMATICAL MODELS OF PROGRESSIVE DISEASES AND SCREENING

BY

ABDOLLAZIM HOSHYAR

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1978

To my wife, Mehri,
and my daughter, Hani,
who have made numerous sacrifices so that
I might pursue this goal.

ACKNOWLEDGMENTS

I wish to express my sincere appreciation and gratitude to the members of my doctoral committee for their overall guidance, understanding and friendship in assisting me in my research. In particular I am deeply indebted to my committee chairman Dr. Ralph W. Swain not only for his technical insight and timely assistance but for suggesting the area of screening for research and for his attitude toward my work. His exemplary contributions to my education will never be forgotten. I wish to extend my appreciation to my committee co-chairman Dr. Thom J. Hodgson for his assistance and encouragement. I would like to thank the other members of the committee, Dr. Kerry E. Kilpatrick, Dr. Gary J. Koehler and Dr. Jeffrey P. Krischer for their comments on drafts of the manuscript. I would also like to thank Dr. Jeffrey P. Krischer and Dr. Lawrence S. Frankel for providing useful references.

I would like to thank Dr. A Ghavami, Deputy Chancellor, Dr. M.S. Mayeri, Dean of the School of Engineering, Dr. M. Bahadori-Nejad and all the faculty members of the department of Mechanical Engineering of Pahlavi University for offering me a scholarship which made my post-graduate study possible.

I am also grateful to the Division of Health Systems Research and the Department of Industrial and Systems Engineering for their financial support during the study.

Finally Mrs. Beth Beville deserves many thanks for her excellent typing of the dissertation.

TABLE OF CONTENTS

		PAGE
ACKNOWLEDGMENTS		iii
KEY TO SYMBOLS		vi
ABSTRACT		ix
CHAPTER		
ONE	INTRODUCTION	1
TWO	LITERATURE REVIEW	5
THREE	MODEL DEVELOPMENT	52
	3.1 A General Model of Cancer Screening	52
	3.2 The Disease Process	54
	3.3 The Screening Process	58
	3.4 Estimation of Transition Probabilities	66
	3.5 Objective Functions	72
	3.6 An Alternative Expression for the Objective Function.	76
FOUR	NEUROBLASTOMA	93
	4.1 Literature on Neuroblastoma	94
	4.2 Development of Unconditional Probabilities	105
	4.3 Objective Function	114
	4.4 Parameter Estimation and Determination of Optimal Policies	116
	4.5 Sensitivity Analysis	121
	4.6 Results and Conclusions	124
FIVE	SPECIAL CASES	133
	5.1 Dependency of True Positive Rate of Two Successive Examinations	134
	5.2 True Positive Rate as a Function of Time from Onset of the Disease	136
	5.3 Screening Examination Consists of a Sequence of Tests	139
	5.4 Transient Problem	154
	5.5 Application of the Model to the Case of Breast Cancer Using the Results of HIP Study	161

CHAPTER PAGE

SIX DETERMINATION OF AN OPTIMAL POLICY 180

6.1 Branch and Bound Method: Search for an Optimal Solution 181
6.2 Search for a "Good" Heuristic Solution 197

SEVEN CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH 205

APPENDIX DERIVATION OF THE GENERAL EXPRESSION FOR TRANSIENT PROBLEM 208

REFERENCES 212

BIOGRAPHICAL SKETCH 220

KEY TO SYMBOLS

$A_1 = 1 - F$

$a_{i'jt}$ = Transition probability of going from i to j in the t - th period without screening

a_{ijt} = Transition probability of going from i to j in the t - th period under any screening program

$B_1 = 1 - F + F/b_{it'}$

$b_{i't}$ = Probability of transition from stage i to i' in the t - th interval without screening

b_{it} = Probability of transition from stage i to i' in the t - th interval under any screening program

$C_d(t)$ = Cost associated with death of an individual of age t

$C_s(t)$ = Screening cost at age t

Csp = Cost of screening the population of suspects to the disease

$C_T(t,i')$ = Treatment cost of an individual who is in stage i' at age t

$C_{t,i'} = C_T(t,i') + C_d(t) \cdot d_i(t)$

$d_{i'}(t)$ = Probability of death due to disease for an individual who is detected at stage i at age t

$f_i(t)$ = Probability that a person who has been in stage i at age t is properly classified as diseased

$f(t)$ = True positive rate of screening as a function of the time from onset

F = Constant true positive rate of screening

I = Number of population per one diseased individual

LB = Lower bound to the objective function

M = Number of screenings

OF = The objective function

O = Occult stages = {1,2, ... N}

D = Detected stages = {1',2', ... N'}

$P_L(t)$ = Probability of not dying of other causes up to age t

$P'[X_t = i]$ = Probability of being in an occult stage i at age t under no screening program

$P[X_t = i]$ = Probability of being in an occult stage i at age t under any screening program

$P'[X_t = i']$ = Probability of being in a detected stage i' at age t under no screening program

$P[X_t = i']$ = Probability of being in a detected stage i' at age t under any screening program

$R(\ell,k)$ = The ratio of population who have completed (ℓ-k) examinations up to ℓ - th interval

r = Discount rate

S = State space (stage of the disease, age of the individual)

T_i = Time of the ith test i = 1, ... M

T_{max} = Maximum age under consideration

$U_{i't}$ = Probability of staying in stage i in the t - th period without screening

U_{it} = Probability of staying in stage i in the t - th period under any screening program

Z_t = Decision variable = $\begin{cases} 0 \text{ if test is not done in the t - th interval} \\ i \text{ if test is done in the t - th interval} \end{cases}$

0 = Healthy state

λ_0 = Rate of onset of the disease

λ' = Rate of detection of the disease

λ_1 = Rate of transition from $(-)_s$ category to $(+-)_s$ category

λ_2 = Rate of transition from $(-)_s$ category to $(++)_s$ category

λ_3 = Rate of transition from $(+-)_s$ category to $(++)_s$ category

$\delta_i = P[X_0 = i]$

Abstract of Dissertation Presented to the Graduate Council
of the University of Florida in Partial Fulfillment of
the Requirements for the Degree of Doctor of Philosophy

MATHEMATICAL MODELS OF PROGRESSIVE DISEASES AND SCREENING

By

Abdollazim Hoshyar

August 1978

Chairman: Ralph W. Swain
Major Department: Industrial and Systems Engineering

A stochastic model for a screening program is presented in which the natural history of the disease is assumed to progress through a set of stages before detection. A model of the whole process is developed which addresses the interaction of the disease process and screening process. The purpose of the model is to develop insight into the disease process and to derive policies which are optimal relative to the particular objectives chosen.

The model is implemented for neuroblastoma and breast cancer and, in the latter case, a comparison is made between the model's output with that of an actual study group. An investigation is also made of the special cases when there is dependency between screening results, when the true positive rate of screening is a function of time from onset, when multiple tests may be employed in each screening and/or when the prevalence pool has a transient period.

To determine the "best" policy, an attempt is made to solve the optimization problem, but to avoid an extensive computation time a

"heuristic" method is also developed which determines "good" policies efficiently. It is shown that a number of screening strategies yield objective function values close to the minimum value for the example studied.

CHAPTER ONE

INTRODUCTION

The fact that some diseases may reach an advanced state without obvious symptoms, coupled with potentially better cure rates associated with early detection, indicates the potential benefits of early detection of the disease. The probability of detection can be increased by the process of screening. While screening programs continue to attract an increasing number of researchers from all fields of science, most of the work done is primarily medically oriented and, generally, there remains a lack of a unifying theory relative to the timing and type of screening used.

The analysis of screening processes centers around determination of an optimal number of examinations at specific ages. This objective is of interest for its theoretical and applied consideration. Theoretically it is a challenging problem which can be difficult to model. Less than perfect predictability, which is a characteristic of screening procedures, plus variability of the characteristics of the disease in different individuals makes the problem difficult to analyze. The real life situations involving different costs and benefits associated with different screening plans provide practical interest in determining good screening strategies. There is general agreement that significant differences exist in the use of one screening policy over another [32,35,49,52,78.89,98,99], and in some cases, such as pap smear and x-ray examinations, the screening itself may not be harmless [35]. Both cost

and health effect factors motivate the need to evaluate alternative screening patterns in a systematic manner.

1.1 Objectives of This Research

A few investigators have presented heuristic procedures for screening of a specified disease. One reason for the lack of a good mathematical model for screening is the lack of understanding of the disease process itself. If a particular disease could be modeled mathematically, with its parameters realistically estimated, then the problem of screening would be easier to analyze. In order to present a "good" screening policy, there is a strong need for a basic mathematical model of the disease.

This research is directed toward the development of optimal screening policies. To accomplish this task for any specific disease it is necessary to decide 1) who should be screened, 2) how often they should be screened and 3) what combination of test methods should be used at each examination. A model of the whole process will be developed, which addresses the interaction of the disease process and screening process. The purpose of the model is to develop insight into the disease process and to derive policies which are optimal relative to the particular objectives chosen. Some of these objectives are: minimization of the total cost of medical care (including the cost of screening and treatment), maximization of life expectancy, maximization of the probability of detection of the disease in a favorable stage and minimization of the delay in time between onset of the disease and its detection.

Increasing the number of examinations increases the screening cost, but hopefully, to some extent, it also increases the chance of earlier detection of the disease. Age of the individual at the time of examina-

tion is a basic factor and depending on the disease under consideration, it can play a major role. Screening at an age in which disease is not likely to be active will increase the cost and may introduce some side effects; screening at the proper age helps the physician detect the disease before its natural detection time and therefore increases the chance of survival and may decrease the cost of treatment. Also using different combinations of screening methods might increase the screening cost, but it decreases the chance of false negative results. Determination of basic factors affecting a screening program, presentation of a mathematical model for that particular disease with estimation of its parameters, validation of the model and selection of the best screening policy depend on the objectives chosen.

Therefore the objective of this research is to develop a general model of cancer screening, specialize it for Neuroblastoma and breast cancer, investigate model sensitivity to changes in parameters or assumptions, and derive good screening policies by use of optimization theory.

Chapter Three provides a general formulation of screening programs and presents a model of the screening process superimposed on the disease process. Restrictive assumptions on the process as well as a discussion of relevant optimality criteria are presented. This chapter is intended to provide a general model applicable to any disease. It consists of four parts: part A presents a stochastic model of disease progress in an individual, part B postulates the probability structure for the progress of the disease and employs available data to estimate the elements of a transition probability matrix, part C presents different

objective functions of interest and estimates the terms, part D contains the basic idea of screening and develops different strategies of screenings.

Chapter Four is based on the development of the previous section and extends the results to the case of Neuroblastoma which is a cancer of early childhood.

In Chapter Five several assumptions of the model are altered and a sensitivity analysis is employed to determine the robustness of the model to changes in some parameters. In the first three sections the assumption of constant screening accuracy is relaxed and the expected total cost of screening is expressed as a function of screening policy for different cases. In the last two sections, the initial effect of a screening program on population prevalence pool is investigated and the model is implemented for the case of breast cancer.

In Chapter Six an attempt is made to solve the optimization problem efficiently and determine the "best" policy. Finally a "heuristic" method is presented which determines "good" policies.

CHAPTER TWO

LITERATURE REVIEW

The problem of screening progressive diseases has received considerable attention in the recent years due to the impact it has on society as a whole. Daniel G. Miller [71] under the topic "What Is Early Diagnosis Doing?" looks at cancer as a very peculiar disease

> . . .Cancer affects the health of other members of society as well as the patient. A disease, the cost of which is fifteen billion dollars a year, one half of the annual budget deficit, has to be considered a matter which affects the health of the nation, if only because it diverts medical manpower and facilities which could be used for other urgent health care needs. Furthermore, a disease which disrupts families,. . . must be considered in terms of its total impact on the society. Currently one in every six dollars spent in health care is spent on cancer, . . .the cost of surgery, radiation therapy, and terminal care is estimated to be at twenty-thousand dollars per case, . . .

Miller points out the importance of early detection and mentions that the use of an appropriate screening policy might reduce the cost associated with cancer.

Gilbertsen [36] in his report on 14,978 cases of cancer claims that his studies suggest that the majority of commonly occuring cancers can be detected on periodic examination far earlier than is likely when the patient waits until his symptoms force him to see the physician and prompt the physician to undertake examination and eventual diagnosis. The result of his studies also suggest that when cancers so detected are treated promptly and adequately, substantial improvement in prognosis

for survival can be anticipated for patients with most of the common cancers which occur today. This concept has been pointed out by many other investigators [4,19,27,45,69,83].

Early diagnosis, which is possible through screening at-risk individuals, is an important concept in aborting or ameliorating the consequences of such diseases as cancers. Considerable funds and effort are being expended for the purpose of the early detection of cancers. The investigations are mostly concerned with the comparison of the costs and benefits of alternative individual screening strategies for selected site-specific cancers and with the selection of preferred strategies. An individual screening strategy is defined [98,99] as the specification of the number and type of screens to be given to an individual in a particular at-risk group and the ages at which the tests are to be given. Screening is operationally defined [99] as the process of selecting those asymptomatic persons who would benefit from further diagnostic studies. The selection of one screening strategy over another depends on the criteria by which the strategy is evaluated and the constraints placed on alternative screening schedule.

Before reviewing the literature a set of definitions related to the screening will be presented.

1) The true positive rate of screening is the probability the test indicates an affected individual, given that the individual has that particular cancer.

2) The false positive rate of screening is the probability the test indicates the disease is present when the disease is not present.

3) The true negative rate of screening is the probability the test indicates the disease is not present when the disease is not present.

4) The false negative rate of screening is the probability the test fails to indicate the presence of the disease in an affected individual.

5) Onset time is the first age at which some recognizable biological change occurs.

6) The pre-clinical state of the disease is regarded as a state where clinical symptoms have not been exhibited and the individual is unaware of the disease.

7) The clinical state of the disease is a state where clinical symptoms are exhibited.

8) Clinical surfacing: If the disease is not detected by a scheduled screening examination, the disease will be said to surface clinically.

9) Lead time for a screening program is the difference between time of diagnosis by the screen and that later time when the disease would be clinically apparent and detectable.

To make a cost-effectiveness analysis, the following requirements would have to be specified by the analyst.

1) The necessary data on the nature of the disease--including detection rates, onset rate, and survival rate.

2) A decision as to what assumptions are to be accepted concerning the disease progression in the occult part.

3) A specification of the accuracy and reliability of the different screening methods, including false-negative rates, false-positive rates and the degree of the uniformity of the test results.

4) A specification of the constraints on the number of screens and their interval per lifetime of an at-risk individual.

5) A specification of the effectiveness measures of interest.

To implement a model to a site-specific cancer there is a strong need for the appropriate data. Unfortunately for most types of cancer, data do not exist and/or are not in a useful form, which makes it extremely difficult to estimate most of the epidemiologic parameters. For data to be useful, it should carry information on the individual's age, sex, race, social level, education, age at each screen, mode of screening, result of screens, whether or not the individual was eventually found with disease through screening, the cause and date of his/her death. Recently there has been some effort to collect data in a more useful manner. For instance, in the case of breast cancer, in December 1963, the Health Insurance Program of Greater New York (HIP) started a long term randomized trial directed at the question "Does periodic breast cancer screening with mammography and clinical examination result in a reduction in mortality from breast cancer in the female population." Two systematic random samples, each consisting of 31,000 women aged 40 to 64 were selected. Each woman in the study group was offered a screening examination and three additional examinations at annual intervals each consisting of a clinical examination, mammography, and an interview. The women in the control group were matched to the study women for date of entry and continued to receive their ordinary medical care [85,86,87,95,101]. These data have been used extensively by many investigators for the purpose of modeling the behavior of breast cancers [39,89,90], estimation of epidemiologic parameters [49,55,56,87], estimation of false negative rates in medical screening [39], estimation of the effect of screening on 5-year mortality rates [85,86,87,95,101]

and determination of factors that correlate with promptness in seeking diagnosis [31].

The accuracy and reliability of the different screening methods are important factors in early detection of the disease and for any particular type of screen an estimate of the false positive rate and false negative rate of screening has to be made.* In 1942 Sawitz and Karpinos [82] proved that when each individual in a group receives the same number of examinations, estimation of the efficiency-true positive rate of screening - $\hat{F}$, and of the prevalence rate, P, are given by

$$\hat{F} = \frac{R[1-(1-\hat{F})^n]}{n\,K} \; ; \quad \hat{P} = \frac{R}{nN\hat{F}} = \frac{K}{N\cdot[1-(1-\hat{F})^n]}$$

where R is the total number of positive test results, n is the number of examinations given to each individual, K is the number of individuals with any positive test results, and N is the number of individuals in the group. In this formulation $\hat{F}$ is the adjusted ratio of positive examinations to the number of examinations on detected diseased individuals and $\hat{P}$ is the adjusted ratio of detected diseased individuals to the number of individuals examined. Therefore knowing R, n, N and K, the procedure can be used to get an estimate of true positive rate of screening and prevalence.

Later, in 1951 Mantel [68] modified these estimates for the case of unequal number of examinations given to different suspects.

---

*Although the literature on estimating the prevalence and true positive rate of screening is cited here, these techniques are not used in the development of the model.

Wittes and Sidel [112] presented a method for estimation of population size from the simple capture-recapture model and showed how to obtain estimates when there are more than two independent sources of notification. They defined

K = the number of notification sources

E_i = probability that a member of the population is identified by i-th source

N = the total size of the population

n_i = the number of members of the population identified by the i-th source

n_{ij} = the total number of members of the population identified by the i-th and j-th source

n = the total number of different population members identified.

They found that $\hat{E}_1$ is the solution of the following polynomial of degree K-1

$$n_1 - n_1 \cdot \prod_{i=1}^{K} (1 - n_i \cdot \hat{E}_1 / n_1) - n \cdot \hat{E}_1 = 0$$

Estimate of $E_2, \ldots, E_n$ are obtained from $\hat{E}_1$

$$\hat{E}_s = \frac{n_s}{n_1} \cdot \hat{E}_1 \qquad s = 2, \ldots, n$$

The population size is estimated as

$$\hat{N} = n / [1 - \prod_{i=1}^{K} (1 - \hat{E}_i)]$$

They gave an approximation to the variance of N and tabulated it for some values of E_1, E_2 and n.

Goldberg and Wittes [39] proposed a similar capture-recapture method [109,110,111] to estimate the sensitivity of the screen, the false negative rate, the population prevalence, the population incidence, their means, variances, and covariances, and their properties in small strata. They used HIP findings [87] to illustrate the model. They defined

i = number of the screens $i=0,\ldots,S$

M_i = number of individuals attending the i-th screen

d_i = number of cases of disease detected at the i-th screen

N_0 = true population prevalence at the 0-th, or initial screen

N_i = true population incidence between the (i-1)-st and the i-th screen ($i>0$)

Θ_i = the number of false negatives

Φ_i = the number of false positives.

Suppressing the subscript i, at any screen $\phi = \frac{\Phi}{M-N}$, $\theta = \frac{\Theta}{N}$ where ϕ and θ are false positive rate and false negative rate. They assumed that false positives have been removed and considered a screening program in which subjects are screened S+1 times by K different screening methods. Let

d_{ij} = number of cases of disease detected at the i-th screen by the j-th screening mechanism. ($i=0, ..S$, $j=1, ..K$) An unbiased estimate of the total number of diseased individuals at the i-th screen, N_i, is

$$\prod_{j=1}^{K} (\hat{N}_i - d_{ij}) = (\hat{N}_i + 1)^{K-1} \cdot (\hat{N}_i - d_i) \qquad i=0,\ldots,S$$

They also computed the variance of $\hat{N}$ for the case K=2.

It is seen that the general limitation of all these procedures is that, for some diseases, the assumption that examination efficiency

is the same for all infected individuals may not be realistic, and instead examination efficiency may vary from individual to individual, depending on the nature and stage of the disease.

It is generally agreed that many cancers can be modeled as progressing through different stages [5]. The staging of the disease is useful, because of its importance in treatment and its effect on prognosis. In December 1965, the American Joint Committee on Cancer Staging [16] offered the TNM system which was based upon the three capital letters; T - Tumor or primary lesion and its extent; N - lymph nodes of the region and their condition; M - distant metastasis. Within each letter element, increasing involvement was categorized by the combination of the capital letter with a numerical suffix. But this staging was too general to be used for all types of cancer. In fact, depending on the type of cancer under consideration different investigators employed different staging methods.

Eker [22] introduced a method of staging for Carcinomas of the colon and rectum in 1963.

Cutler and Myers [17] introduced a method of staging for breast cancer in 1967.

Barron and Richart [6] introduced a method of staging for the cancer of cervical carcinoma in 1968.

Evans et al. [27] introduced a method of staging for Neuroblastoma in 1971.

Aside from the fact that the importance of early detection is known, there are only a few mathematical models for screening, none of which is general enough to cover all forms of the disease. In what follows, the literature on screening models is reviewed and general concepts of

screening programs as a tool for early diagnosis of malignant diseases are pointed out.

In 1963 Lincoln and Weiss [65] derived properties of the time with disease before detection in recurrent screening and investigated the consequences of these properties with respect to the interval between screens. They considered the efficiency of different policies for scheduling medical examinations, and treated both periodic and random examinations allowing for imperfect diagnosis depending on how long the disease had been present. Working with that portion of the population in which a tumor appears, they defined

$U(t)$ = Probability density for the time at which one can first observe the presence of a tumor.

$\alpha(t)$ = Probability that a diagnosis made at a time t after the appearance of the first observable signs of tumor, will be incorrect.

Let examinations occur at times $\tau_1, \tau_2, \ldots$, such that the intervals $\Delta_1 = \tau_1, \Delta_2 = \tau_2 - \tau_1, \ldots$ are independent identically distributed random variables with probability density function $\phi(\Delta)$. Then under the assumption that a tumor is discovered only by examination, they used the idea that the examination times $\{\tau_i\}$ form a renewal process, and if the tumor first becomes observable at t, the time to detection is

$$T_d = T_1^* + T_2 + T_3 + \ldots + T_n$$

where n is the number of tests before detection, T_1^* is the forward delay, or time to first test following the initiation of the disease, and T_j is the time between the j-1st and jth test. Then

$$\eta(t,x) = [1-\alpha(x)]f(t,x)$$

where $\eta(t,x)$ = probability density for the time to discovery of the tumor conditional on its having become observable at time t,

and $f(t,x)$ = probability density for the event that a test occurs at t and that any diagnostic made in $(t,t+x)$ were incorrect.

Also,

$$\eta(x) = \int_0^\infty u(t)\eta(t,x)dt$$

where $\eta(x)$ = probability density of tumor age at discovery.

They found the expression for moments of $\eta(x)$ as functions of $\alpha(t)$ and $\Phi(\Delta)$. Using the quantities shown and employing the following two optimality criteria they developed an optimal schedule:

Criterion 1: No more than a fraction $\varepsilon<1$ of those people who eventually have a tumor will have an undetected tumor for more than a specified time T.

Criterion 2: Mean undetected time of tumor growth does not exceed a given time T_e.

Weiss and Lincoln [106] used the model in [65] for the case of cervical cancer. They used a gamma distribution for u(t) and a negative exponential for Δ_i where $\Delta_i = \tau_i - \tau_{i-1}$. Neglecting death from other causes, they obtained some characteristics for screening period by using an (a,b) policy, where a is time and b is a probability. The (a,b) policy is one in which the probability that any tumor is of age a or older at the time of discovery is b.

In 1967 Feinleib [29] presented the mathematical justification and restrictions for the well-known epidemiologic relation that the prevalence of a disease is proportional to its incidence and mean duration. Let P(t)

be the prevalence of a disease at time t; i(t) the incidence of disease at t; and g(d|t) the conditional probability density function of durations of incident cases where d is the duration from time of onset, then if P(0) = 0,

$$P(t) = \int_0^t \int_{t-y}^{\infty} i(y) \cdot g(x|y)\, dxdy$$

For the stable disease model, he imposed the following three restrictions

1) $i(t)=i$ for all $t \geq 0$ and is zero for $t < 0$.

2) $g(d|t)=g(d)$ for all $t \geq 0$.

3) $g(d)=0$ for $d > M$.

For this model he proved that $P(t) = i \cdot \bar{d}$ for $t > M$, where $\bar{d}$ is the mean duration.

In 1968 Hutchison and Shapiro [49] published their work on the estimation of some parameters of preclinical breast cancer. They used preliminary findings of a clinical program of screening for breast cancer to estimate average duration of preclinical disease (early stages of the disease, in which tumor may be detected only in a screening program). They assumed that in any large population there are some individuals with preclinical disease. The number of such women (prevalence) depends on the rate with which new preclinical disease develops (incidence) and the length of time (duration) it persists before clinical diagnosis. Then if the number of preclinical cases remains constant during the long run, the rate of new cases must be the same as the rate at which old cases are passing over to clinical disease. In general a prevalence P and an incidence I imply that the average duration is $d = \frac{P}{I}$.

Moreover if, in the absence of screening, the duration of preclinical disease were the same for all individuals, then duration-to-date of those detected at screening would be uniformly distributed between (0,t), average duration-to-date would be half the total duration, and this would be equal to average lead time. In their mathematical model, the only input was

I_d = incidence of cases of duration d. I_d is expressed as a discrete distribution function with $\sum_{d=0}^{\infty} I_d = 1$.

Where d = duration, or interval of time during which an individual case is diagnosable by screening but not diagnosed under usual practice. Then they gave functional form for prevalence, incidence, mean duration and lead time as a function of I_d.

For instance,

$$P_n = \sum_{d=0}^{n} I_d \cdot d + n \cdot \sum_{d=n+1}^{\infty} I_d$$

where P_n = prevalence at time n of cases detectable by screening but not diagnosed under usual practice.

Also

$$I_{(n)} = \sum_{d=0}^{n} I_d(n-d)$$

where $I_{(n)}$ = total incidence in interval n following screening.

Using preliminary HIP findings [84,85], they estimated the average duration of preclinical breast cancer in absence of special screening to be 20 months. This will give an average lead time of 10 months for a

completely homogenous population. Later on it will be seen that there are some assumptions inherent in this model which makes their analysis restricted.

In 1969 Blumenson and Bross [11] presented a mathematical analysis of the growth and spread of breast cancer. They developed a mathematical model which describes the development of the cancer from the appearance of the first cell to the possible occurrence of a distant metastasis. Their model also accounts for limitation on the minimum size of the tumor before it can be detected and for the effect of surgical intervention by the physician on the development of a recurrence of the disease. They used a deductive method and constrained the progress of breast cancer to the contribution of the following parameters (1) the tumor doubling time, (2) patient's delay in reporting her disease to the physician, (3) the chance that the disease will spread to nearby lymph nodes, and (4) the chance of spreading to more distant parts of the body. A patient is classified as having a small primary tumor (S) - average diameter less than 5 centimeters, or a large primary tumor (L). A patient either has negative nodes (0), 1-3 positive nodes (2), or more than three positive nodes (4). After surgery the patients are followed for at least 18 months. At the end of this period a patient is classified as (N) or (R) depending on whether she had no clinically detectable recurrence. This introduces 12 stages; S0R, S2R, S4R, ... They presented a method for calculating the probability of being in any of these states as a function of the parameters introduced. Once the twelve probabilities have been calculated, they are compared with the data and a χ^2-test is used to determine those values of the parameters which minimize the χ^2-value.

Calling their model a deep mathematical model* for human breast cancer, interestingly, they found out [14] that the only way to get an acceptable χ^2-value is to employ a two-disease hypothesis for breast cancer, i.e., the population of patients consists of two groups with different rate of doubling time.

In 1969 Zelen and Feinleib [116] presented a model of a chronic disease which progresses from a pre-clinical state to a clinical state and related the potential benefit of the screening program to the lead time gained by early diagnosis. They developed a stochastic model for early detection programs which led to an estimate of the mean lead time as a function of observable variables. They considered a screening program where an individual was examined only once. In their model, transitions are from a disease-free state (S_0) to a preclinical disease state (S_p), and then to a clinical disease state (S_c). They assume that (S_p) eventually progresses to (S_c) if not detected and treated. They define

$q(t)$ = p.d.f. of sojourn time in S_p

$Q(t) = \int_t^\infty q(x)\, dx$

$P(t)$ = probability of being in S_p at time t

$Q_f(t)$ = unconditional forward recurrence time (lead time) distribution

m = mean sojourn time in $S_p = \int_0^\infty Q(x)\, dx$

---

*A deep model describes an underlying process which in theory, generates the surface events.

and assume that

$$Q_f(t) = \frac{\int_t^\infty Q(y)\, dy}{m}$$

$$L = \text{mean lead time} = \frac{m^2+\sigma^2}{2m} = \frac{m}{2}(1+C^2)$$

where $C = \frac{\sigma}{m}$, and m and σ^2 are the mean and variance of the sojourn time distribution in S_p. They mentioned that $L > \frac{m}{2}$ is because of the length-biased sampling which means the screen does not detect people at random, but detects people with longer preclinical sojourn times. They also mentioned that particular care must be exercised if one is comparing the survival of the individuals detected early by a test with a comparable group of individuals detected at a clinical state because those found in S_p tend to have longer preclinical sojourn times in S_p than the general population. This might be synonomous with a slow growing disease in the preclinical state. Zelen and Feinleib consider different conditions under which the following relationship referred to by Hutchison and Shapiro [49] and Feinleib [29] is valid:

$$P = m \cdot I$$

where P, I and m are prevalence, incidence and mean duration of the disease. They proved that even if prevalence and incidence are time dependent, P(t)/I(t) is equal to the mean sojourn time in S_p provided the sojourn time follows an exponential distribution. In general

$$I(t) = \frac{P(t)}{m} + \tfrac{1}{2}(C^2-1)\, P'(t) + \ldots \qquad \text{where } C = \frac{\sigma}{m}$$

They developed relationships among age, prevalence and incidence and employing data from HIP [87] found an estimate of 1.84 years for m.

Finally they relaxed the assumption that every individual eventually leaves the (S_p) and developed some relationships which depend on the ratio of those who eventually will surface to the total population and is a generalization of their model to non-progressive diseases.

In 1972 Kodlin [57] used a series of biometric arguments and very simple cost estimates to attack the cost-benefit problem in screening for breast cancer and found that survival results would justify the increased costs that might result from mass screening. To do this he claimed that in the case of breast cancer, one is faced with essentially two alternative basic strategies:

1) To screen and bring the positives to therapeutic intervention--associated total cost is called C_1.
2) To let the cases come to diagnosis and treatment through the traditional pathway of recognition--associated total cost is called C_2.

Let t = treatment cost for those picked up by screen

t' = treatment cost for those not picked up by screen

W = W_0 (physical exam) + (biopsy fee) W_1

S = screen cost

b = biopsy rate amongst false positives

Then

$$C_1 = \overbrace{P \cdot \Pi_1 \cdot (t+W+s)}^{\text{true positive}} + \overbrace{P(1-\Pi_1) \cdot (t'+W+s)}^{\text{false negative}} +$$

$$\overbrace{(1-P) \cdot (1-\Pi_2) \cdot (W_0+b \cdot W_1+S)}^{\text{false positive}} + \overbrace{(1-P) \cdot (\Pi_2) \cdot S}^{\text{true negative}}$$

where P = presumed frequency of breast cancer in the population

Π_1 = conditional probability of identifying a case correctly by mammography and palpation

Π_2 = conditional probability of identifying a non-case correctly.

Also

$$C_2 = P\cdot(t'+W) + \phi\cdot(1-P)\cdot[m+(1-\Pi)\cdot(W_0+b\cdot W_1)]$$

where

ϕ = fraction of the non-diseased who demand a breast check-up, and

m = mammography cost.

Using some estimate of the parameters P, Π_1, Π_2, ϕ and costs, he found that C_1 is usually greater than C_2. Then he attempted to assess the total cost per case cured and using this objective found that it would be beneficial to choose the first strategy, screening.

Shapiro, Goldberg and Hutchison [84] used the experience in the HIP Study [87] to estimate the average time gained through screening in the detection of breast cancer among the women who were aged 40-64 years at the start of the screening program. They used the model presented by Hutchison and Shapiro [49] which says d = P/I and found a mean duration of 1.3 years for a prevalence of 2.73 and incidence of 2.09 as calculated from HIP data. The statistical models that were applied suggested that the average lead time was about a year.

In 1974 Kirch and Klein [53] developed methods for determining the optimal screening policy using the criterion of detection delay which is time from first point at which disease is detectable to the point of actual detection. They started with an age-dependent disease and showed that, under certain conditions in an optimal schedule, the interval between examinations is proportional to the square root of the age-specific incidence of the disease. They were interested in possible

advantages of nonperiodic policies in mass screening, and tried to find out whether a nonperiodic schedule, involving the same expected tests per patient as a periodic schedule, could reduce the average time to detect a given disease, or, whether a nonperiodic schedule involving fewer expected tests per patient could lead to detection of the disease as early as a given periodic schedule.

They divided population into two groups, those who will eventually get the disease and those who will not. For the first group, interest was centered on early detection. For the second group, interest was centered on minimization of the expected number of tests, and found a class of optimal schedules by varying the number of screenings. They assumed

1) The age span consists of "equal length periods" in which incidence rate is usually tabulated in the literature.
2) Each such period starts with an examination and all examinations within a period are at equal intervals.
3) Examinations are error-free.

Define: T = Earliest time at which the disease could be detected, if an examination took place (they assumed T is uniformly distributed over the interval).

x_i = Number of tests scheduled for i-th period.

D = Length of time between T and examination time.

$Q(x_i,D)$ = Detection delay if the disease becomes detectable in the i-th period.

S_i = Probability that patient survives to the start of period i.

Assuming that T has a uniform distribution, they found the expression for $E[Q(x_i,D)]$ as a function of D and x_i for two different cases:

1. D is a constant.
2. D is a random variable with probability density function $P_D(t)$.

They used the optimization model

$$\text{Minimize} \quad G(x_1,\ldots,x_n;D) = \sum_{i=1}^{n} P_i \cdot E\,[Q(x_i,D)]$$

$$\text{S.T.} \quad \sum_{i=1}^{n} x_i \cdot S_i \leq K \qquad (K \text{ is a constant} \geq n)$$

$$x_j \geq 1$$

where P_i represents the conditional probability that the detectability point occurs in period i, given that it will occur sometime within the n periods of interest.

Kirch and Klein [52] applied this model to breast cancer and found that there is a slight economic advantage (2% to 3%) if examination frequency is taken to be a function of age rather than fixed throughout life. Their model has some basic assumptions (such as screening is error-free) which are not always true, and consequently restrict the application of their model.

In 1974 Tallis and Sarfaty [97] presented a model for the distribution of the time to reporting cancer, called T. The basic assumption of their model was that the infinitesimal conditional probability of reporting the disease is proportional to the rate of tumor growth. Let the distribution function of T be $F(t) = P[T \leq t]$ with derivative $F'(t) = f(t)$, then the force of reporting function is

$$\frac{f(t)}{1-F(t)} = \phi(t) = \begin{cases} 0 & 0 \leq t < t_0 \\ \alpha \cdot V'(t) & t \geq t_0 \end{cases}$$

where $V(t)$ is the tumor volume t time units after onset, α is a constant of proportionality and t_0 is defined by the equation $V(t_0) = v_0$. For instance, v_0 may be the minimum clinically detectable tumor size. Then

$$F(t) = \begin{cases} 0 & 0 \leq t < t_0 \\ 1-\exp(-\alpha \cdot V(t)) & t \geq t_0 \end{cases}$$

They defined $R=T-t_0 | T \geq t_0$, and used a staging method similar to TNM [16] which consisted of four stages S_1, S_2, S_3 and S_4 in such a way that each stage is associated with certain volumes of tumor growth, i.e., $v_{i-1} \leq V < v_i$ is classified as S_i, $i=1,\ldots 4$.

Let r_i be such that $V(t_0+r_i)=v_i$, $i=0,1,2,3$, $(r_i=0)$, then if P_i is the probability of reporting the disease in S_i,

$$P_1 = P[R < r_1] = 1-\exp(-\alpha \cdot v_1+\alpha \cdot v_0)$$

$$P_2 = P[r_1 \leq R < r_2] = [\exp(-\alpha \cdot v_1)-\exp(-\alpha \cdot v_2)]\exp(\alpha \cdot v_0)$$

$$P_3 = P[r_2 \leq R < r_3] = [\exp(-\alpha \cdot v_2)-\exp(-\alpha \cdot v_3)]\exp(\alpha \cdot v_0)$$

$$P_4 = P[r_3 \leq R] = \exp[-\alpha \cdot v_3+\alpha \cdot v_0]$$

They used a standard model for tumor growth $V(t_0+r)=v_0 \cdot e^{\beta r}$ where β is the rate of growth and found $E(R)$ and $E[V(R)]$ as functions of α and β,

$$E(R) = e^{\alpha \cdot v_0} \beta^{-1} \int_{\alpha \cdot v_0}^{\infty} y^{-1} e^{-y} \cdot dy$$

$E[V(R)]$ = expected tumor volume at reporting

$$= \frac{1+\alpha \cdot v_0}{\alpha} \simeq \alpha^{-1}$$

Knox [55,56] in his simulation studies of breast cancer screening programs postulated a model for the natural history of the disease which consisted of a set of stages. A statement of the natural history of the disease was then provided in the form of a transition matrix which gave estimated transfer rates between the various stages. This set of values was adjusted iteratively until an output was provided which matched available data on incidence, prevalence and mortality. He used HIP findings as his data [87].

The full set of stages for the simulation included 26 defined stages and the general sequence of the transfer pattern was held constant. The simulation was developed in four steps.

1) Some specifications for the natural history of the disease were developed which make the model capable of explaining available incidence, prevalence, mortality and case-fatality data.

2) Sensitivities and specificities for palpation and mammography were developed that make the model capable of explaining the results of HIP.

3) The results were extrapolated to predict the benefits of extended screening services.

4) The scope of Urinary-Steroid prescreening tests was examined.

He points out that mortality savings are not to be seen as the sole criterion and it is possible to provide short but useful prolongations of life without grossly affecting the cumulative death statistics. He also mentions that

> . . .results of the HIP experiment are not the results that we shall be needing. We would need to know quite accurately, for each age group, the effectiveness, hazards

> and costs of an extended palpation program using staff other than doctors, together with the teaching of self-palpation, and the marginal benefits, risks and costs of a selective and limited use of mammography, superimposed upon this background. . . . Having regard to the limited acceptability of screening procedures, their high costs, . . ., a reasonable service target may be a reduction of breast cancer mortality by about one-tenth.

In 1976, a significant amount of research was completed. Prorok [78,79] presented a stochastic model for a periodic screening policy in the case of a chronic disease whose natural history is assumed to follow a progressive path from a preclinical state to a clinical state. The distribution of the forward recurrence time (time interval from initiation of the disease to its detection) is derived and used to obtain the distribution and mean of the lead time (time interval between the point at which early diagnosis occurs as a result of a screening test and the point when disease would have been detected in the absence of screening), and the relationships for calculating the proportion of preclinical cases detected. Prorok defines (S_0) to be disease-free state, (S_p) preclinical state and (S_c) the clinical disease state. He develops a model for the interaction of the disease process with an independent screening process. Assuming that screening time is independent of the time at which the preclinical state is entered and the duration of stay therein, and allowing for the possibility of imperfect detection, he derives all necessary distributions.

Prorok also measures the number or fraction of preclinical cases which are actually detected earlier than usual as a result of screening. His model is a generalization of the model presented by Zelen and Feinleib [116] for the case of multiple screening. The restriction on his model and models similar to it is due to the method of staging

used, and the assumption of periodic screening. The staging (S_0, S_p, S_c) is very general in nature and by using it, we lose the information available from data. Periodic screening is also a restriction which is not necessarily optimal, since there are some indications that, in the case of age-dependent diseases, the optimal screening policy would not be periodic [53].

Galliher [35] has raised the question of how soon a repeat pap smear should be scheduled in the case of a woman who has had no test or has received negative smears up to date in her life. He used cost-effectiveness analysis to determine optimal frequencies of such schedules. Normally, there is a trade off between the amount of effort at prevention and the resulting amount of advanced disease if not screened. To analyze this problem there is a strong need for data on the occult part of the disease, but, in almost all cases, the early disease has always been treated when detected. Therefore, due to the lack of appropriate data, he bounded the objective by obtaining two extreme sets of objectives, between which the true evaluation of objective should lie. He introduced two extreme levels, (L) and (M). Under L one assumes: 1) carcinoma-in-situ could come to detection between the scheduled preventive smears in more than 50% of all affected individuals; 2) only 40% or so of all carcinoma-in-situ would ever progress to invasive disease if not treated. Therefore under (L) one obtains a minimal role to the periodic screening. Under M one assumes: 1) carcinoma-in-situ would never come to detection except at the scheduled preventive examinations; 2) carcinoma-in-situ, if untreated, would always eventually progress to invasive cancer, and that invasive cancer is always preceeded by carcinoma-in-situ. Therefore (M) provides

a maximal role to the periodic screening. In his paper, Galliher only worked out level (M), because it offered certain simplicities in the task of producing a cost-effectiveness analysis.

Galliher assumed a course based upon four possible stages of diagnosis at screening, and defined them as

H = Healthy

U = Carcinoma-in-situ

V-C = (Micro) invasive disease, curable at detection

V-F = (Micro) invasive disease, fatal even though treated.

Figure 1 is a flow chart for the onset and course of the disease in an individual.

Galliher's main assumptions are

1) Carcinoma-in-situ does not regress spontaneously.

2) Duration from onset of carcinoma-in-situ to onset of invasion is negative exponentially distributed (he used a mean duration of ten years in his computations).

3) Carcinoma-in-situ alone will not be detected by discomfort of the woman.

4) Carcinoma-in-situ is completely curable at diagnosis.

5) There is a sharp transition to invasion at the end of the duration of carcinoma-in-situ.

6) When the stage of invasion is entered, two competing processes commence and operate concurrently; these are

a) tendency to surface clinically,

b) tendency to make the transition to fatal disease.

7) The rate for transition to fatal disease and clinical surfacing are constants g and c, independent of how long invasion has been in

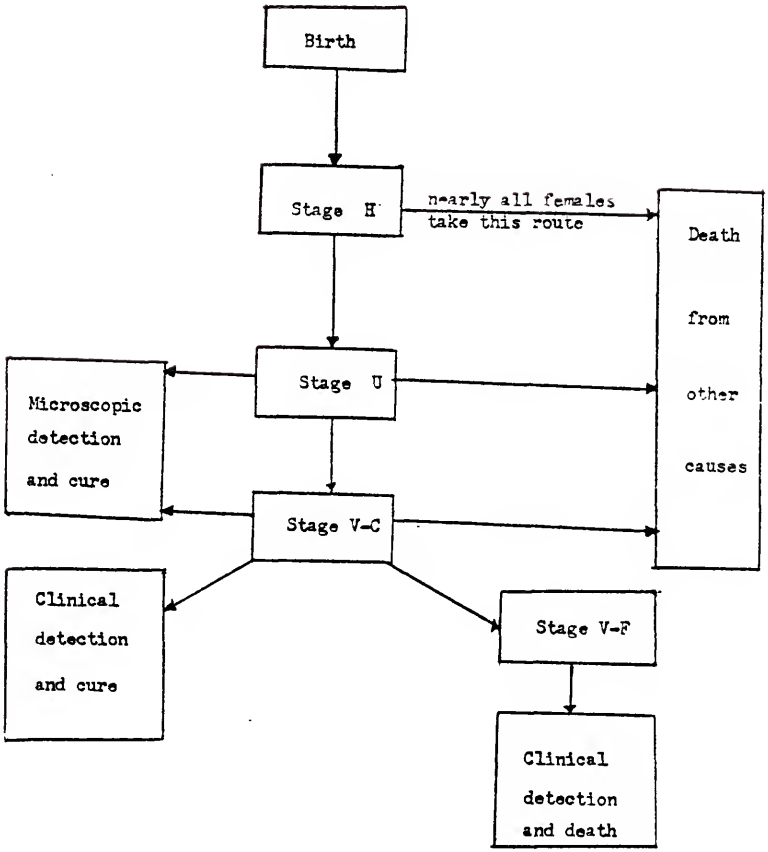


Figure 1: Flow chart for onset and course of the disease.

process. Therefore if the invasive condition is not first diagnosed, the chance that it remains curable for at least the next t years is equal to e^{-gt}, and the chance that the invasive condition remains clinically occult for at least t years is equal to e^{-ct}.

8) Initially, the individual starts at 20 years old and has no risk for previous cervical cancer.

He employed the following measures of effectiveness in his computations.

1) The minimization of the total medical care.

2) The minimization of sum of 1) plus the economic loss to the society if the patient dies.

3) The minimization of premature death.

Galliher used data to estimate incidence of onset, g, c and different elements of cost. For the numerical task of finding optimal schedules he needed the probabilities

P(i, j, m, t) = The probability that an individual who is in stage i of the disease at the end of the m-th time interval will be in stage j of the disease at age t if she does not have a pap smear between the two stages. This quantity could be easily computed from the above mentioned assumptions.

A mathematical model of breast cancer which is similar to the work of Galliher [35] was developed by Shwartz [88,89,90,91,92,93]. His model was based upon the hypothesis that breast cancer is a "time dependent" disease, and if it is detected and treated earlier, the prognosis will be more favorable. However, due to the potential high cost of screening (in terms of dollars, psychological effects, and physical side-effects), it is important to establish estimates of the benefits of screening so

that one might better evaluate if the benefits appear worth the costs.

Literature on breast cancer shows [88] that tumor size and extent of lymph node involvement are the most important prognostic variables that affect prognosis. Employing this, Shwartz assumed tumor growth rate is exponentially distributed, and the rate of lymph node involvement is a function of the size of the tumor and its growth rate. Using these functional forms, he compared the results with available data and computed the parameters. Therefore his model consisted of a set of hypotheses on the incidence of the disease, its progression, its tendency to be detected without benefits of any scheduled screening examinations and the relationship between stage of the disease and survival. Using this model he found, for a woman in a given at-risk level, the level of effectiveness as a function of the number of tests, the corresponding ages at examinations, the reliability of the tests, and whether or not the individual had performed self examinations between scheduled screening examinations. Modifications in the model were made and a second model was proposed so that the two models bracketed relevant hypotheses about the rate of disease progression. The two models differed in their hypothesis about the rate at which lymph nodes become involved.

Due to the lack of appropriate data, Shwartz vaired the assumptions over the entire range and computed the benefits, with the hope that actual process lies somewhere in between. For instance, he parameterized the distribution of tumor growth rates, threat of death from breast cancer, the false negative rate and the correlation between the false negative probability on successive examinations.

Using heuristic techniques rather than optimization techniques, he determined the best ages at which screening examinations should be given to each individual.

The three benefit measures employed by Shwartz were

1) the life expectancy of a women from her current age,

2) the probability that detection occurs before nodal involvement, and

3) the probability that there is no recurrence of breast cancer.

D.E. Thompson and T.C. Doyle [99] proposed an approach to the selection of screening policy for cancer of the colon and rectum. They presented an approach to analyzing the question of how often a person should be screened for colorectal cancer to achieve a desired cost-benefit outcome. To do this they reviewed and analyzed data on the incidence and prevalence of the disease, the course of the disease, the cost and ability of detecting the disease at various stages of progression, the cost and effectiveness of treating the disease in any stage and benefits of treatment.

Thompson and Doyle offered two approaches to modeling the disease.

1) A continuous model, based on tumor size and growth data.

2) A discrete state model, based on progress of the disease through several stages.

In this paper they emphasized the second model, which, in the absence of data reflecting the true progress of the disease, has been structured in the context of available data and is relatively simple in design. This simplified version of the discrete model was used to perform parametric analysis.

Their model has two principal functions:

1) It provides a basis for determining the relative values of different screening policies.

2) It provides a means of analyzing the sensitivity of screening policies to variations of the onset and progress of the disease, the efficiency of screens, costs associated and/or benefits.

Their discrete model can be viewed as consisting of

1) A set of discrete states, and

2) A set of probability distributions, which describe the length of time that an individual remains in a particular state.

The model consisted of the following stages:

H : Health

A : Lesion confined to the mucosa but undetected

A': Lesion confined to the mucosa and detected

B : Cancer into the muscularis propria with negative lymph nodes but undetected

B': Cancer into the muscularis propria with negative lymph nodes and detected

C : Cancer with nodal involvement but undetected

C': Cancer with nodal involvement and detected

D : Distant metastases present but undetected

D': Distant metastases present but detected

M : Death.

Due to the unavailability of appropriate data, they proposed a particular staging and obtained few primary conclusions. In this model the status of an individual is related to whether disease, when detected and treated, is curable or not, i.e.,

H : Health

P : Occult Colorectal Cancer, pre-fatal, i.e., curable if detected

P': Colorectal Cancer, detected by clinical surfacing or screening, curable

F : Occult Colorectal Cancer, fatal

F': Colorectal Cancer, detected and fatal

M : Death.

They assumed that the death rate ($\mu(a)$), onset rate ($P(a)$), rate of progression from pre-fatal to fatal stages (λ), rate of clinical surfacing of pre-fatal (ψ_p), rate of clinical surfacing of fatal colorectal cancer (ψ_f) and rate of death associated with detected fatal cancer (μ') were constants (i.e., they are assumed to have negative exponential distribution). This assumption resulted in a Markov process:

Let $P_i(t)$ = the probability that an individual is in stage i at age t.

Then

$$P_i(t+s) = \sum_j P_j(t) \cdot B_{ji}(t,t+s)$$

or

$$\underline{P}(t+s) = \underline{P}(t) \cdot B(t,t+s)$$

where $B_{ji}(t,t+s)$ = probability of a change from j to i in the time interval from t to t+s.

Then screening was introduced into the model by means of the probability of detecting occult cancer, i.e.,

a_{k_0} = P{Detection of occult cancer, given that the disease is in stage k}

They considered a screen at time t_1, and denoted the time immediately after the screen by t_1^+; then

$$P_{i_D}(t_1^+) = P_{i_0}(t_1) \cdot a_{i_0} + P_{i_D}(t_1)$$

$$P_{i_0}(t_1^+) = P_{i_0}(t_1) \cdot (1-a_{i_0})$$

where i_0, i_D correspond to occult and detected stages of the disease. Then

$$P(t_1^+) = P(t_1) \cdot A$$

where

$$A_{ij} = \begin{cases} 1 & \text{for } i=j \text{ and } i \text{ corresponds to a detected stage of the disease} \\ 1-a_i & \text{for } i=j \text{ and } i \text{ corresponds to an occult stage of the disease} \\ a_i & i \text{ and } j \text{ correspond respectively to occult and detected stages of the disease} \\ 0 & \text{otherwise} \end{cases}$$

Therefore for time t_2

$$\underline{P}(t_2) = \underline{P}(t_1^+) \cdot B(t_1,t_2)$$

$$= \underline{P}(t_1) \cdot A \cdot B(t_1,t_2)$$

and so on.

They examined their model using the following three measures:

1) Oncological: More effective strategies are characterized by porportionately higher detection in earlier stages.

2) Medical costs: This includes the cost of the screening program itself, the cost of diagnosis associated with false positives, the

cost of treatment and the costs attributed to the period of disability of a patient.

3) Life expectancy: More effective strategies are characterized by longer expected length of time prior to death.

D.E. Thompson and R. Disney [98] introduced a general mathematical model of progressive diseases and screening. Conceptually, the disease history could be represented as progressing through a series of stages whose durations are random variables and whose meaning could be interpreted in terms of the individual's prognosis and the sensitivity of the disease to detection in that stage through clinical surfacing and/or screening. Their model was a mathematical model of the interaction of two independent random processes, namely, the disease process and the screening process. The purpose of the disease portion of the model was to predict a person's status at any age t. The model can compute the probability that an individual is healthy or in some stage of the disease and, if he has had the disease, the length of time he had been in the particular stage.

They assumed that the screening methods could produce false negative or correct results, and the probability of these results could depend on the stage of the disease, the length of time in the current stage and the particular method used for screening. In their model, as soon as an individual dies of other causes or is detected as having the disease, the process terminates. Moreover, the time from birth to death from other causes is assumed to be a random variable independent of the disease process. This assumption makes the analysis much simpler, because it makes it possible to model the disease portion as a semi-Markov process

whose transition matrix is independent of the age of the person. This assumption allows them to use the existing theory of Semi-Markov processes.

Let $E = \{0,1,2,\ldots,S,S+1,\ldots,N\}$ be the state space of the disease.

State 0 indicates absence of disease, and state 1 through S indicate that the individual is in one of the S occult stages of the disease. Stages S+1 through N indicate that the disease has surfaced clinically, or been detected by screening. Let Y_t be the state of the individual, given that he is alive at age t.

T_n = age of the individual when he changes stage for the n-th time

X_n = Length of time that an individual will spend in the state he entered when he was T_{n-1} years old.

They assume that everybody starts life free of disease. This is not generally true (for example in the case of Neuroblastoma), and the analysis can be modified to take care of this possibility.

Thompson and Disney propose the following assumption:

$$P\left\{Y_n = j, X_n \leq x \,|\, Y_{n-1} = i, Y_{n-2}, \ldots, Y_0, X_{n-1}, \ldots, X_0\right\}$$
$$= P\left\{Y_n = j, X_n \leq x \,|\, Y_{n-1} = i\right\} = A_{ij}(x)$$

where

$$P(Y_0=j) = \begin{cases} 1 & \text{if } j=0 \\ 0 & \text{if } j\neq 0 \end{cases}$$

$$P(X_0=0) = 1$$

and compute the probability of the event

$$B = \{Z_t : Y_t = j,\ U_t > x,\ V_t > y\}$$

where V_t and U_t are foreward and backward recurrence times of the process.*

They define quantity $Q_t(j,x)$ to be: $Q_t(j,x) = P\{Y_t = j,\ U_t \leq x\}$, and consider the screen given at the time $\hat{T}_m = t$. Assuming $Q_t^+(j,x)$ denotes the distribution of (y_t,U_t) immediately after this screen, they evaluate $Q_t^+(j,x)$ as a function of $Q_t(j,x)$. So given the initial condition $(Y_0 = 0,\ U_0 = 0)$ at birth, the equation $P\{Y_t = j,\ U_t \leq y | Y_s = i,\ U_s = x\}$ can take the distribution of (Y_t,U_t) up to the time of the first screen at age $\hat{T}_1$. Then $Q_t^+(j,x) = f[Q_t(j,x)]$ modifies this distribution at the time of the first screen, as individuals are taken from the occult part of the disease to the detected states. A similar analysis is applicable for the interval $\hat{T}_1$ to $\hat{T}_2$ and to the screen given at age $\hat{T}_2$, etc.

This model is restricted to the case of diseases which are stationary in time and are not age dependent. Unfortunately, the literature on most of the cancers shows that there is a significant difference in relative survival statistics among young and old individuals [5]. Therefore, in those cases this model has to be modified to take care of time dependence of the disease.

---

*The forward recurrence time at age t is defined as the interval from age t to the epoch of the next state change of $\{Y_t\}$, that is $V_t = T_n - t$, if $T_{n-1} \leq t < T_n$. Similarly, the backward recurrence time at age t is $U_t = t - T_{n-1}$ if $T_{n-1} \leq t < T_n$.

Albert and Louis [2,3,66] have done very broad research on the subject of screening of progressive diseases in the last two years. In their first paper [2] on screening for the early detection of cancer, they have characterized the natural history of a chronic disease state in terms of the distribution of X (a person's age at the time of entering the disease state), Y (the sojourn time in that disease state), and A (a person's present age) over a population of individuals. Then, they have defined age specific incidence, prevalence, life time attack rate, mean duration of the disease state, cohort effect, etc., in terms of the joint distribution (X, Y, A), which for known (X, Y, A) gives a method for estimation of those parameters. In their second and third papers [3,66] they have found the impact of screening on the natural history of the disease and presented a method for estimation of the disease natural history. A brief review of their work follows.
They define

1) $f_{XYA}(.,.,.;t)$ = joint distribution for (X,Y,A) at any instant t.

2) A person is nonsusceptible if $X = \infty$.

3) A person is a chronic habitue of S if, for that person $X < \infty$ and $Y = \infty$.

$\therefore$ proportion of chronic habitue = $Pr.\{X < \infty, Y = \infty\}$

4) A cohort effect is said to exist if the distribution of (X,Y) varies over age strata.

5) The lifetime attack rate of disease state S is

$P[X < \infty]$;

$$\therefore \quad P[X < \infty] = P[x < \infty, Y = \infty] + \int_0^\infty \int_0^\infty \int_0^\infty f_{XYA}(x,y,a)\, dx\, dy\, da$$

6) $I_S(a)$ = The age specific incidence of S among those aged a:

$$\therefore \quad I_S(a) = f_{X,A}(a,a)/f_A(a) = f_{X|A}(a|a)$$

7) I_S = The overall incidence of S:

$$\therefore \quad I_S = \int_0^\infty I_S(a) \cdot f_A(a)\, da$$

8) $\phi_S(a)$ = The age specific prevalence of the disease state among those aged a.

$$\therefore \quad \phi_S(a) = \int_{x=0}^{a} \int_{y=a-x}^{\infty} f_{XY/A}(x,y/a)\, dy\, dx$$

9) ϕ_S = The overall prevalence of S:

$$\therefore \quad \phi_S = \int_0^\infty \phi_S(a) \cdot f_A(a)\, da$$

10) If there is no cohort effect and if $E[X|X<\infty]$ is finite, then

$$E[Y|X<\infty] = \frac{\int_0^\infty \phi_S(a)\, da}{\int_0^\infty I_S(a)\, da}$$

This equation is a generalization of d=P/I used by other investigators [29,49,116]. They define S_1 to be the state which is entered upon leaving S, and show that $I_{S_1} = I_S$, which is used by other researchers, is true if the age distribution is flat and there is no chronic habitue.

They also present a method for construction of (X,Y) distribution from prevalence and incidence information.

If there is no cohort effect

$$f_X(x) = I_{S_1}(x) + \frac{d}{dx}\phi_S(x)$$

If there is no cohort effect and X and Y are independent, then

$$\bar{f}_Y(S) = \frac{\bar{I}_{S_1}(s)}{[s\cdot\bar{\phi}_S(s)+\bar{I}_{S_1}(s)]}$$

where $\bar{f}$, $\bar{I}$ and $\bar{\phi}$ are laplace transforms.

Then using counter examples they prove that without the assumption of independence (between X and Y), the Y distribution is not unique.

Since Y denotes the preclinical latency and X_0 denotes the age at time of entering the disease state, then at the instant of time X_0+Y the patient surfaces. Now, consider a population of individuals at a certain instant of time, t, each person has an associated vector of sojourn time $\underline{Z} = (X_0, X_1, \ldots, X_k, Y)$. This plus the person's age, A(t), describe the natural course of the disease in that individual. Denote the density of $(\underline{Z}, A(t))$ over the study population $f_{\underline{Z},A}(t)$ and the total population size by N(t).

In the absence of screening, they allow people to leave the study population for two reasons:

a) Death from competing risk

b) By reason of surfacing with the type of cancer under study

and assume that

a) A clinically surfaced individual leaves the study population forever,

b) False positives are eventually discovered and returned to the study population. They actually take this probability to be zero.

Then if $n^{(t)}(\underline{Z},a)\cdot dA\cdot d\underline{Z}$ is the number of individuals in the population at time t that occupy the cell $(\underline{Z},\underline{Z}+d\underline{Z})\cdot(a,a+da)$, they show that

$$\frac{\partial n^{(t)}}{\partial t}(\underline{Z},a+t) = M(t)\cdot \tilde{f}^{(t)}_{\underline{Z},A}(\underline{Z},a+t) - \psi^{(t)}(\underline{Z},a+t)\cdot n^{(t)}(\underline{Z},a+t)$$

provided $0 < a+t < X_0+Y$

and of course in the complementary region, $n^{(t)}(\underline{Z},a+t) = 0$.

Here

$M(t)$ is the immigration rate,

$\tilde{f}^{(t)}_{\underline{Z},A}(\cdot)$ is the joint density of $\underline{Z}$ and A among those who immigrate in at time t,

and

$$\psi^{(t)}(\underline{Z},a) = r(\underline{Z},a,t)\cdot\phi(\underline{Z},a) + d(\underline{Z},a,t)$$

where

$r(\underline{z},a,t)$ is the screening rate at time t in the stratum $\underline{Z}=\underline{z}$, $A(t)=a$, $\phi(\underline{z},a)$ is the probability of a positive screen if $\underline{Z}=\underline{z}$ and $A(t)=a$, and $d(\underline{Z},a,t)$ is the death rate at time t in the stratum $\underline{Z}=\underline{z}$ and $A(t)=a$. Therefore $\psi^{(t)}(\underline{Z},a)$ is the instantaneous net rate of removal of individuals from the above mentioned stratum.

The solution to the above differential equation is

$$n^{(t)}(\underline{z},a) = \begin{cases} [n^{(0)}(\underline{z},a-t)+K^{(t)}(\underline{z},a-t)]\cdot\exp[-Q^{(t)}(\underline{z},a-t)] \\ 0 \quad \text{otherwise} \end{cases} \qquad \text{if } 0 < a < X_0+Y$$

where

$$Q^{(t)}(\underline{z},a) = \int_0^t \psi^{(u)}(\underline{z},a+u)\,du$$

and

$$K^{(t)}(\underline{z},a) = \int_0^t M(v)\cdot \tilde{f}^{(v)}_{\underline{Z},A}(\underline{z},a+v)\cdot\exp Q^{(v)}(\underline{z},a)\,dv$$

To find the joint density of Z and A in the study population at time t, $n^{(t)}(\underline{z},a)$ should be normalized

$$f^{(t)}_{\underline{Z},A}(\underline{z},a) = \frac{n^{(t)}(\underline{z},a)}{\iint n^{(t)}(\underline{z},a)\, d\underline{z}\, da}$$

Therefore the effect of screening on $f^{(t)}_{\underline{Z},A}(\cdot)$ can be computed which gives a means to predict the temporal behavior of epidemiologic parameters in the presence of screening.

In order to answer the question of "what is a better strategy of screening?" they introduce the concept of "critical point." Treatment that is begun before this time point has a relatively high probability of success, whereas treatment begun after this point has a markedly lower probability of success. Their objectives are

1) The discovery of rate of less-favorable-prognosis disease, $1f(t)$; $1f(t)\,dt$ is expected number of cases with less favorable prognosis that are diagnosed in the interval $(t, t+dt)$.

2) The salvage rate $s(t)$; $s(t)\,dt$ is the expected number of cases discovered by screening in $(t, t+dt)$, who have favorable prognosis, but have $Y > X_1$.

The following data are required as input to their computations.

1) Age specific death rate $d(a,t)$

2) Immigration rate, $M(t)$

3) Initial population size, $N(0)$

4) Age specific screening rate, $r(a,t)$, as a function of time

5) The screening detectability function, $\phi(\underline{z},a)$

6) The initial distribution for (X_0, X_1, Y, A)--$f^{(0)}_{\underline{Z},A}(\underline{z},a)$.

7) The (possibly time varying) distribution for (X_0, X_1, Y, A) among immigrants--$f^{(t)}_{\underline{Z},A}(\underline{z},a)$.

In their third paper [66] Louis, Albert and Heghinian present a non-parametric method for estimation of $f_{XYA}(.,.,.:t)$. To observe the accuracy of their estimates, they generated data, used it to find the estimates of certain epidemiologic parameters and compared these estimated values with their corresponding theoretical values and the usual epidemiologic estimates. Several points should be noted about their work:

1) The method of staging employed is very restrictive and the fact that there is a unique passage from S_0 to S_1.. to S_k makes it impossible to employ their model to the case of diseases such as Neuroblastoma where from S_3, the individual can either go to S_4 or S_5.

2) Their objective function is not strong enough to use all information on survival. In fact it picks a point--critical point t_c--and weights every point $t < t_c$ with value one and every point $t > t_c$ with value zero.

3) The effect associated with immigrates is not well defined and since the distribution is assumed to be known, in most cases it decreases the accuracy. It would be much better to neglect its effect than to bring it in at an unknown level of effectiveness. They mention that 20% immigration did not change the screening policy significantly. This is why other investigators neglect its effect and assume a closed population.

Having reviewed the literature on the topic of screening processes, we observe that the main issues on this complicated subject are

1) An understanding of the natural behavior of the disease

2) Presentation of a method for estimation of the disease parameters.

3) Investigation of the impacts of screening procedures on performance measures of interest such as the probability of detection, the probability of recurrence and the probability of survival (in an individual).

4) Investigation of the impacts of screening on the whole society.

No single model has been able to answer all these issues rigorously. Some unanswered questions are related to the last subject. Most investigators have tried to find the benefits of lead time on the total population. It has been claimed that those who are detected by the screening procedure have a longer preclinical time and are considered to be slow developing disease individuals. Therefore any measurement solely based on this group would be biased and determination of the benefit of lead time for the general population would be imperfect.

From a modeling point of view, the most general models are those of Lincoln and Weiss [65,106], Klein and Kirch [52,53], Zelen and Feinleib [29,113,114,115,116], Bross and Blumenson [10,11,14,15], Prorok [78,79], Galliher [35], Shwartz [88,89,90,91,92,93], Thompson, Disney and Doyle [98,99] and Albert [2,3,66]. These models cover a variety of methods, some of which are similar in the direction of their conclusions but some are not. The structure of the model developed in this research is similar to the models developed by Albert, Thompson and Disney, Shwartz and Galliher but the specific formulation and general conclusions differ. The following pages summarize the material review in this chapter.

No	Researcher	Disease	Goal of the model	Basic assumptions	Conclusions
1	Sawitz & Karpinos	general	Estimation of the true positive incidence rate	Every individual participates in all scheduled exams	They presented an expression for estimation of the true positive incidence rate
2	Mantel	general	Estimation of the true positive incidence rate	Relax the above assumption	They presented an expression for estimation of the true positive incidence rate
3	Furnival	Breast cancer	Evaluate the usefulness of existing screening methods	---	Neither thermography nor mammography is effective as an isolated screening procedure
4	Goldberg & Witts	general-applied to breast cancer	Estimation of the sensitivity and specificity of the screen. The population prevalence and incidence	False positives are removed	They presented an expression for the unbiased estimate of total number of diseased individuals at any test
5	Feinleib "1967"	general	Relationship between prevalence, incidence and mean duration	-Incidence of the disease is invariant of time-p.d.f. of duration of incidence cases is not a function of time	Under those two assumptions, prevalence=incidence·mean duration

No	Researcher	Disease	Goal of the model	Basic assumptions	Conclusions
6	Hutchison & Shapiro "1968"	Breast cancer	Estimation of the average duration of preclinical disease	Average duration = prevalence/incidence	Using HIP data, they estimated the average duration of the breast cancer to be 20 months
7	Blumenson & Bross "1969"	Breast cancer	-Presentation of a mathematical model of the growth and spread of the disease -Comparison with data to estimate the value of parameters	Effective elements in progress of breast cancer: i) tumor doubling time ii) delay in reporting the disease iii) chance of lymph node involvement iv) chance of spreading to more distant parts	The population of the patients consists of two groups with different rate of tumor growth
8	Zelen & Feinleib "1969"	general -applied to breast cancer	A stochastic model for early detection programs which considers the potential benefit of a screening program as a function of the lead time gained by early diagnosis	1) Every individual is screened only once 2) Every individual is surfaced eventually 3) Disease progresses from a preclinical state to a clinical state	The screen does not detect people at random, but detects people with longer preclinical sojourn time
9	Kodlin "1972"	Breast cancer	Cost analysis of breast cancer screening programs	He makes a series of biometrical arguments and very crude cost estimates	Survival results would justify the increased cost associated with mass screening

No	Researcher	Disease	Goal of the model	Basic assumptions	Conclusions
10	Shapiro Goldberg & Hutchison "1974"	Breast cancer	Use the HIP data to estimate the average time gained through screening	They employed model #6.	Average lead time is about a year
11	Tallis & Sarfaty "1974"	general -applied to breast cancer	Determination of the distribution of the time to reporting cancer	1) Conditional probability of reporting the disease is proportional to the rate of tumor growth 2) Distribution of tumor growth is exponential	They presented an expression for expected time to reporting the cancer and its variance
12	Knox "1974"	Breast cancer	-Estimation of the parameters of the disease using simulation -Extrapolation to determine the results of any screening policy	Disease progresses through a set of stages	HIP results are not the results that are needed
13	Prorok "1976"	general	-A stochastic model for a periodic screening policy. -Derivation of the distribution and mean of the lead time	-Staging consists of a preclinical and a clinical state -Screening is done periodically	His model is a generalization of model #8 for the case of multiple screening

No	Researcher	Disease	Goal of the model	Basic assumptions	Conclusions
14	Thompson & Disney "1976"	general	A general mathematical model of progressive disease and screening which consists of the interaction of two independent random processes--the disease and the screening process	1) The disease progresses through a series of stages 2) The disease is modeled as a semi-Markov process whose transition matrix is independent of the age 3) Every individual is disease-free at birth	Under those assumptions the probability that an individual is healthy or in some stage of the disease and, if he has the disease, the length of time he has been in that stage could be computed

No	Researcher	Disease	Goal of the model	Basic assumptions	Objectives of interest	Conclusions
15	Lincoln & Weiss "1963"	General -applied to breast cancer	-Derive properties of the time with disease before detection in recurrent screening -Investigate the consequences of these properties with respect to the interval between screens	They 1) treated both periodic and random examination 2) allowed for imperfect diagnosis	1) No more than ε of those people who eventually will have an undetected tumor for more than a specific time T 2) Mean undetected time of the tumor growth does not exceed a given time T_e	Different incidence curves do not produce a significant deviation from a periodic examination policy.
16	Kirch & Klein "1974"	General -applied to breast cancer	Mathematical model for determining examination schedules for the detection of an age dependent disease	Examinations are error-free	Minimize detection delay which is time from onset of the disease to her actual detection	In an optimal schedule, the interval between tests is proportional to the square root of the age-specific incidence rate.
17	Gallither "1976"	Carcinoma in-situ	Presentation of a cost effectiveness analysis to determine optimal frequencies of pap smear	He introduced two sets of assumptions between which he believed the actual case lies. He used method of staging. Carcinoma in-situ does not regress spontaneously.	1) The minimization of the total medical care. 2) The minimization of sum of 1) plus the economic loss to the society if the patient dies. 3) The minimization of premature death.	In most cases, the most benefit is obtained when about 10 pap smears are obtained during life at the proper ages.
18	Shwartz "1976"	Breast cancer	Presentation of a stochastic model for the natural history of the disease. Determination of the optimal screening policy.	1) Tumor size and extent of lymph node involvement are the main variables that effect prognosis. 2) Tumor growth rate is exponentially distributed and rate of lymph node involvement is a function of the size of the tumor and its growth rate.	1) The life expectancy of a woman from her current age. 2) The probability that detection occurs before nodal involvement. 3) The probability that there is no recurrence of breast cancer.	He found the level of effectiveness as a function of different elements.

No	Researcher	Disease	Goal of the model	Basic assumptions	Objectives of interest	Conclusions
19	Thompson & Doyle "1976"	Colorectal cancer	Presentation of a discrete stochastic model based on progress of the disease through a set of stages which provide a basic for determining the relative values of different screening policies.	Death rate, onset rate, rate of progression from prefatal to fatal stages, rate of clinical surfacing and rate of death asso-ted with detected fatal cancer are constants.	1) More effective strategies are characterized by pro-portionately higher detection in earlier stages, 2) Minimization of medical cost, 3) Maximization of life expectancy.	Due to lack of data, there is a need to gather relevant knowledge about this disease.
20	Albert & Louis "1976"	General -applied to breast cancer	-To present a basic pro-babilistic formulation for the natural history of disease in a popula-tion through use of a simple "disease state" model. -A mathematical model of screening -Statistical method for estimating components of the disease natural history.	The joint distribution of current age, age at time of entry into disease state and sojourn time in disease state is the main element of interest.	1) The discovery of rate of less-favorable prognosis disease, 2) The salvage rate, expected number of cases discovered by screening who have favorable prognosis, but if left would have surfaced with un-favorable prognosis.	-Screening has a transient effect on the population. -The model output is only as realistic as the input.

CHAPTER THREE

MODEL DEVELOPMENT

In this chapter, a general model of cancer screening will be developed which incorporates the stochastic nature of the disease process with the viewpoint common to most of the cancer literature — that cancer disease process can be represented as progressing through a series of stages whose durations are random variables.

3.1 A General Model of Cancer Screening

The literature review reveals that almost any site specific cancer could be represented as a process that progresses through a series of stages [16,27,35,85]. The staging of the disease can be done in alternative ways, and there has been a continual argument concerning the choice of one method over another. Staging of cancer found its initial importance in reporting the end results. The "American Joint Committee on Cancer Staging And End Results Reporting" was organized on January 9, 1959, to develop a system of clinical staging of cancer by sites, acceptable to the American medical profession. The committee has completed and published different brochures for clinical staging of different site specific cancers. The main objective for clinical staging, to be useful, is that it should be relatively simple and should yield meaningful information as to prognosis. In the case of modeling a specific cancer, the choice of one method of staging over another depends on

1) The availability of needed information. There are cases where a staging method seems to be easy to apply, but there are no data to support it. In such a case, data must be gathered by employing the staging method under consideration which needs time and the cooperation of different investigative groups.

2) The Complexity of staging versus usefulness of the results. It might be possible to combine data in different manners and design different staging methods, in which case there is a trade-off between complexity of the staging, accuracy of the result and its usefulness.

In the model developed, the disease process is looked upon as if it passes through a series of stages before the individual is clinically surfaced by coming to medical attention through signs and symptoms. At any time, the disease process may be terminated by death from cancer or from causes other than cancer. Throughout this research, attention will be restricted to the disease process and the results will be conditioned on the individual's surviving to the age of interest. The model makes it possible to predict an individual's status at any desired age. It will be possible to compute the probability that a person is free of disease and, if he has the disease, the stage he might be in.

Then the interaction of the disease process with a random process called the screening process will be considered. The screening process is a process in which the individual who is suspected of having the disease is examined by one or more of several screening methods at specific points in time. Based on the result of the screening procedure, the person under consideration will be categorized as "free of disease" or "having the disease." In the first case he will not be considered until the next scheduled examination time. In the second case more

specific examinations are performed to verify initial results and identify the stage of disease. To allow for all possible situations, screening methods are permitted to produce false negative and false positive results. Under a false negative result, the individual is classified as healthy, whereas in reality he is in some stage of the disease. Under a false positive result, the individual is classified as being in some stage of the disease, whereas he is healthy. This introduces some extra costs, due to additional examinations necessary, but eventually it will confirm that the individual is healthy. In the development of the model it will be assumed that the screened population is closed. There is no allowance for emigration from the population under consideration, and death from other causes is independent of the death from cancer. A competing risk approach can be used to tie these two processes together.

3.2 The Disease Process

Data analysis [5,9] shows that for several cancer sites of interest in this research the relative survival statistics are strongly age-dependent and that the distribution of the time from detection of the disease to death from the disease depends upon the individual's age at the time the disease is detected. Therefore, the age of an individual is incorporated as a dimension in the state space of the model. The basic elements of the model are the age of the individual and the stage of the disease he is in.

Suppose that an individual of age t can be in one of $(2N+3) < \infty$ states with respect to a given disease. Let

$$S = \{[0_t, 1_t, \ldots, N_t, 1'_t, 2'_t, \ldots, N'_t]_{t\in\{T\}}, 0', D\}$$

$$T = 0,1,2,\ldots,T_{max}$$

be the state space of the disease, where T is the set of discrete times which are the ages of interest in ascending order and T_{max} is the maximum age under consideration. Depending on the nature of the disease, t could be weeks, months, quarters or years and T_{max} should be large enough to cover screening of any individual carrying the disease.

State 0_t is assumed to indicate that the individual of age t is free of disease. States 1_t through N_t indicate that the individual of age t is in one of N occult stages of the disease. Stages $1'_t$ through N'_t indicate that the disease has surfaced clinically or been detected by examination for an individual of age t, when he was in one of stages 1 through N, respectively. States 0' and D indicate that the individual has been "cured" or "died," respectively. In Figure 2, the stages of the disease are depicted schematically.

An individual may start life healthy and remain in that state until he dies of other causes. or he might start life healthy and develop the disease sometime during his course of life. Then he will be classified as in one of the stages 1 through N. The disease may progress from one state to another until he is surfaced clinically (through signs or symptoms) or is detected by screening. Then he goes from state n ($1 \leq n \leq N$) to state n'. In this simplified model, regardless of the progression of the disease in the detected stages, the individual is assumed to die from the disease or be cured. That is, depending on the age of the individual and the stage he was detected in, whether or not the disease progresses through the detected states, he has a certain chance of going to states 0' and D. In Figure 3, the transitions that are permitted in the model are illustrated in more detail.

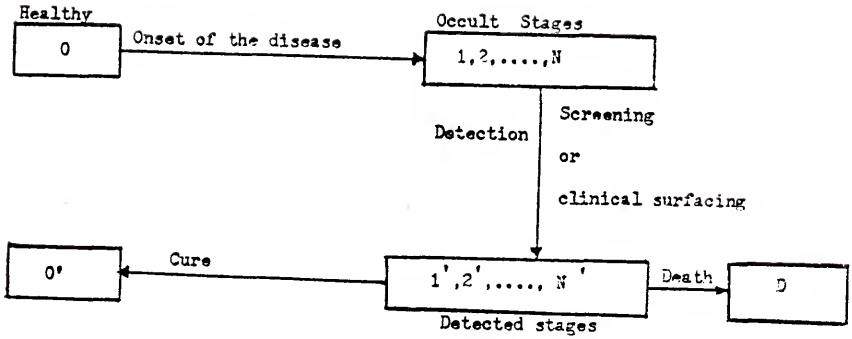


Figure 2: A schematic diagram of the stages of the disease.

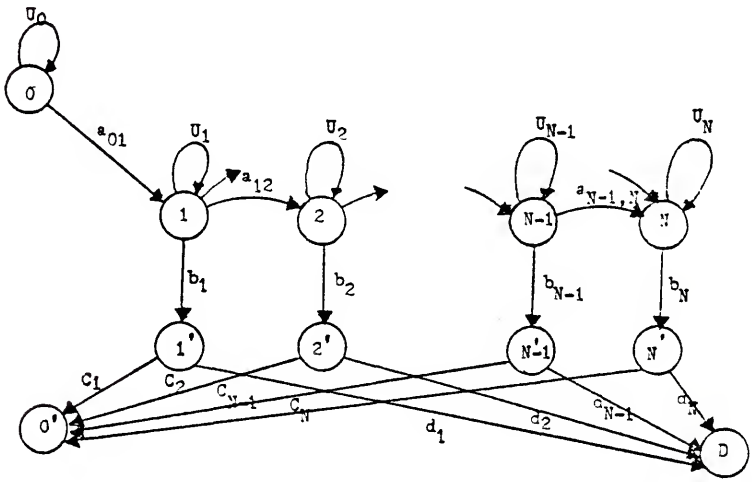


Figure 3: Structure of the disease.

There is no restriction on the condition at birth, which means that a newborn infant could be healthy (with respect to the special disease under consideration) or possibly in any one of the occult stages. This modification plus "time dependency" of the disease would be necessary in the case of some kinds of site specific cancers such as Neuroblastoma.

At any time the disease process can be terminated by death from causes other than the disease, therefore the attention is restricted to the disease process and the results are conditioned on the individual surviving to the age of interest. For modeling purposes, the possibility of death from the disease prior to recognition of the disease is included in the model with clinical surfacing from the disease. This is because death from the cancers is unlikely prior to clinical surfacing of the disease. Therefore, there is no one step transition from occult stages of the disease to the state "death from the disease." The process terminates when the individual dies of any cause or is cured through the process of treatment. The re-examination of the individual after the treatment is done will not be considered. This is because the risk of a second cancer is usually high enough that screening of former patients cannot be considered routine screening.

Let X_t by the state of the individual of age t, given that he is alive. For instance 2_{10} would mean a 10 periods (weeks, months,...) old person is in stage 2 of the disease. Let the disease process be tracked only at the end of fixed intervals of time. Conceptually, the random process $\{X\}$ can be pictured as a process that goes from one state to another. As soon as a new state is reached, the process randomly chooses the next stage to be visited. After choosing this next stage

and depending on which stage the process is currently in and will next go in, the process randomly chooses a time required to make that transition. Therefore the resulting process $\{X\}$ is a Semi-Markov process and the sequence $\{X_n\}$ would be a Markov chain discrete time process. The one-step transition probability

$$P_{ij} = P\{X_{n+1} = j | X_n = i\}$$

has a very special structure, which consists of upper diagonal blocks. This can be seen from Figure 4.

It is seen that the only possible transition for an individual of age t is to go to age t+1, regardless of progression of the disease. To clarify the process, a sample function of the disease process is shown, in which the individual is detected in stage 2 at age 6 and gets cured. See Figure 5.

3.3 The Screening Process

The disease process goes from one stage to another until it is clinically surfaced through the nature of the disease itself. If the disease is a time-dependent process, it usually surfaces at a stage in which there is a relatively poor chance of survival. Therefore, there is an interest in detecting the disease in its early stages of progress, which can be done through screening process. The screening process is a process in which, at specific points in time, the individual who is suspected to have the disease, is examined by one or more of several screening methods. In order to screen for a disease, one needs a simple, rapid, relatively accurate test to select from a general population those persons who would benefit from further diagnostic studies [45]. A screening method has several characteristics, which in

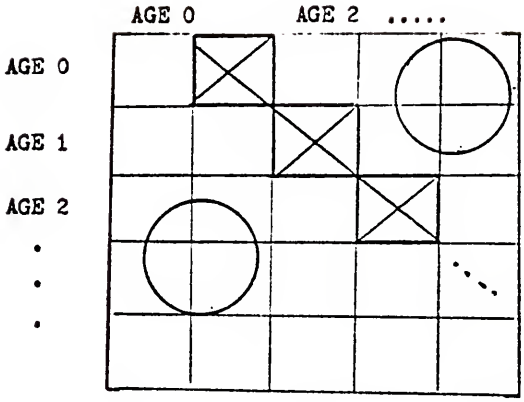


Figure 4: One step transition probability diagram.

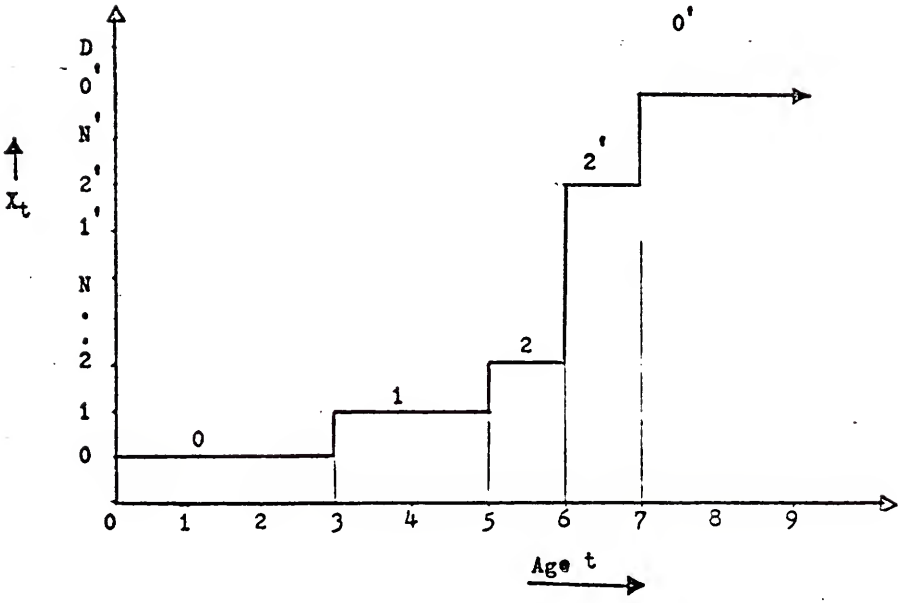


Figure 5: A sample function of the disease process.

general make it possible to choose one method over another, if any one is to be chosen. For instance, a screening method may be harmful [89] because of the unreliability of the technique used, resulting in increased psychological trauma because of high false-positive rates and perhaps decreased surveillance and an unwarranted sense of security because of the high false-negative rates and because of possible deleterious side-effects of screening. In general an ideal screening test must have an acceptable false-negative rate and false-positive rate.

An individual should be screened under a screening policy which determines the ages at which the individual is to be screened, and the method which is going to be used for screening. Assume that there are M different examination methods and that screening examinations are given at ages $T_1 < T_2 < T_3 \ldots$. For modeling purposes assume that no more than one screening test per each time interval is done and the screening examination for the period (t,t+1) is done at t^+, which is a time very close to t but greater than t. A sample function of the process with and without screening is shown in Figure 6.

It is assumed that when an individual in state i ($1 \leq i \leq N$) is screened he will stay in the state i (if test gives a false-negative result) or go to state i' (if the test gives a correct result), and therefore there would be no error in stage recognition of the screening method. In the following material, the effect of screening on the process will be determined.

Let b_{it} = Probability of transition from stage i to i' in the t-th interval with screening.

b'_{it} = Probability of transition from stage i to i' in the t-th interval without screening.

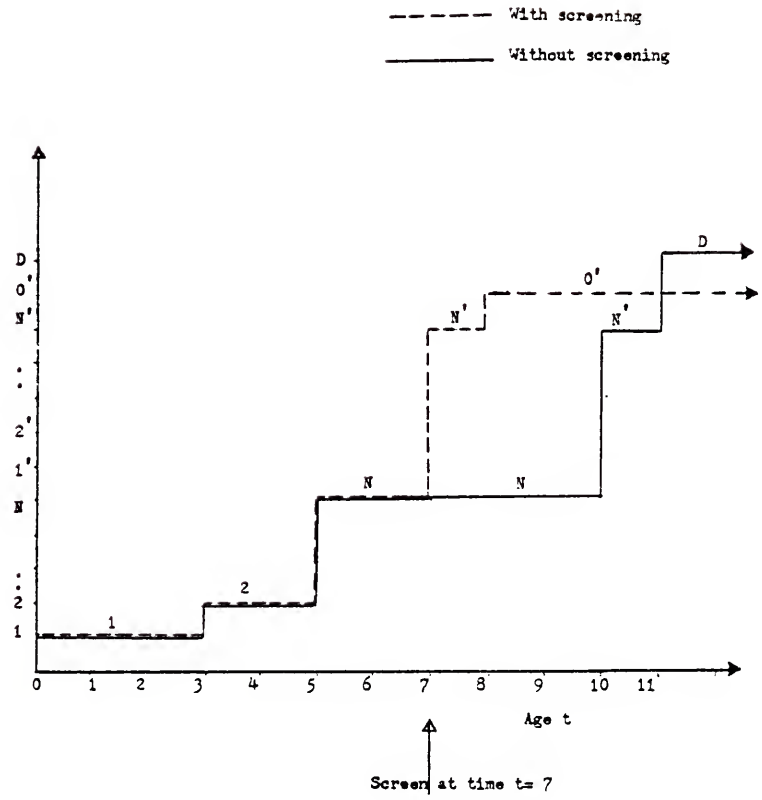


Figure 6: A sample function of the process with and without screening.

Define $f_i(t)$ to be the probability that a person who has been in state i at age t is properly classified as diseased. Then screening will affect b_{it} in the following manner:

$$b_{it} = P[\text{test is done in the t-th interval}] \cdot P[\text{individual being in state i at the time of test}] \cdot P[\text{test gives correct result}] + P[\text{test is done in the t-th interval}] \cdot P[\text{test gives false-negative result}] \cdot b'_{it} + P[\text{test is not done in the t-th interval}] \cdot b'_{it}.$$

Define Z_t as a zero-one variable, which takes value one if the test is done in the t-th interval and zero otherwise, i.e.,

$$Z_t = \begin{cases} 0 & \text{If test is not done in the t-th interval} \\ 1 & \text{If test is done in the t-th interval.} \end{cases}$$

Then

$$b_{it} = P[Z_t \neq 0] \cdot f_i(t) \cdot P[X_{t+}=i \mid X_t=i] + \left\{P[Z_t=0] + [P[Z_t \neq 0] \cdot (1-f_i(t))]\right\} \cdot b'_{it}$$

$$b_{it} = Z_t \cdot f_i(t) \cdot P[X_{t+}=i \mid X_t=i] + \left\{(1-Z_t) + Z_t \cdot [1-f_i(t)]\right\} \cdot b'_{it}$$

$$b_{it} = Z_t \cdot f_i(t) \cdot P[X_{t+}=i \mid X_t=i] + [1-Z_t \cdot f_i(t)] \cdot b'_{it}$$

Later, the system will be assumed to be discrete so that the process can jump only at the end of a discrete interval. This assumption is equivalent to saying

$$P[X_{t+} = i \mid X_t = i] = 1$$

in which case b_{it} reduces further to

$$b_{it} = Z_t \cdot f_i(t) + [1-Z_t \cdot f_i(t)] \cdot b'_{it}$$

This means

$$b_{it} = \begin{cases} b'_{it} & \text{If screening is not done in the } (t,t+1) \text{ interval} \\ f_i(t)+[1-f_i(t)]\cdot b'_{it} & \text{If screening is done in the } (t,t+1) \text{ interval} \end{cases}$$

A simple way to see this is by the use of the following argument

$$b_{it} = \begin{cases} b'_{it} & \text{If test is not done} \\ 1 & \text{If test is done and gives true-positive result} \\ b'_{it} & \text{If test is done but gives false-negative result.} \end{cases}$$

But the following probabilities are associated with the above events:

P[test not done in the $(t,t+1)$ interval] $= 1 - Z_t$

P[test is done in the $(t,t+1)$ interval and gives true-positive result] $= Z_t \cdot f_i(t)$

P[test is done in the $(t,t+1)$ interval but gives false-negative result] $= Z_t[1-f_i(t)]$

Therefore

$$b_{it} = \begin{cases} b'_{it} & \text{with probability } (1-Z_t) \\ 1 & \text{with probability } Z_t \cdot f_i(t) \\ b'_{it} & \text{with probability } Z_t \cdot [1-f_i(t)] \end{cases}$$

$$b_{it} = b'_{it} \cdot (1-Z_t) + 1 \cdot Z_t \cdot f_i(t) + b'_{it} \cdot Z_t[1-f_i(t)]$$

$$\Rightarrow \quad b_{it} = Z_t \cdot f_i(t) + [1-Z_t \cdot f_i(t)] \cdot b'_{it}$$

A similar argument can be employed in the case of other elements of the transition probability matrix. Define

a'_{ijt} = Transition probability of going from i to j in the t-th period without screening

a_{ijt} = Transition probability of going from i to j in the t-th period.

Then

$$a_{ijt} = a'_{ijt}\{\text{probability that the test is not done}\} + a'_{ijt}\{\text{probability that test is done}\} \cdot \{\text{probability that test gives false-negative result}\}$$

$$a_{ijt} = a'_{ijt} \cdot \left\{P[Z_t=0] + P[Z_t\neq 0] \cdot [1-f_i(t)]\right\}$$

$$a_{ijt} = a'_{ijt} \cdot \left\{(1-Z_t) + Z_t[1-f_i(t)]\right\}$$

$$a_{ijt} = a'_{ijt} \cdot [1-Z_t \cdot f_i(t)]$$

A simple way to see this is by the use of the following argument:

$$a_{ijt} = \begin{cases} a'_{ijt} & \text{If the test is not done in the } (t,t+1) \text{ interval} \\ 0 & \text{If the test is done in the } (t,t+1) \text{ interval and gives true-positive result} \\ a'_{ijt} & \text{If the test is done in the } (t,t+1) \text{ interval but gives false-negative result.} \end{cases}$$

Therefore

$$a_{ijt} = \begin{cases} a'_{ijt} & \text{with probability } (1-Z_t) \\ 0 & \text{with probability } Z_t \cdot f_i(t) \\ a'_{ijt} & \text{with probability } Z_t \cdot [1-f_i(t)] \end{cases}$$

This gives

$$a_{ijt} = a'_{ijt}[1-Z_t \cdot f_i(t)]$$

Let U'_{it} = Probability of staying in state i in the t-th period without screening

U_{it} = Probability of staying in state i in the t-th period.

The same analysis reveals that

$$U_{it} = U'_{it} \cdot [1-Z_t \cdot f_i(t)]$$

This analysis shows that all elements of the one-step transition probability are affected by screening in the following manner:

Probability element	Notation	No screening in (t,t+1)th int.	Screening done in (t,t+1)th int.
Probability of detection	b_{it}	b'_{it}	$f_i(t)+[1-f_i(t)]\cdot b'_{it}$
Probability of stay	U_{it}*	U'_{it}	$U'_{it}[1-f_i(t)]$
Probability of jump	a_{ijt}*	a'_{ijt}	$a'_{ijt}[1-f_i(t)]$

*Note that U_{ot} and a_{olt}, which are related to the "Healthy" state, are not dependent on the screening procedure and regardless of the screening method used, they will remain unchanged.

It is seen that U'_{it} and a'_{ijt} are decreased and b'_{it} is increased through the process of screening. Later on, in this research, this analysis will be employed to develop a general form for the probability transition matrix.

3.4 Estimation of Transition Probabilities

Any model has certain associated parameters, the realistic estimation of which should be a primary goal of the analyst. In the model employed in this research there are four types of parameters.

1) Probability of staying in the occult stage i in the interval (t,t+1).
2) Probability of going from stage i to j (in the occult part of the disease) in the interval (t,t+1).
3) Probability of detection of an individual in state i at age t.
4) Probability of survival of an individual in state i at age t.

The probability of going from i to j includes onset of the disease as a special case. In order to have a realistic estimate of these parameters, data available in the literature should be employed consistently. The survival probabilities are given in the literature on cancer and a realistic estimate of these probabilities for each stage and age group is not that difficult (although lead time effects are difficult to estimate), but the first three sets of parameters mentioned above are extremely difficult to estimate. This is due to the fact that there is no consistent data on the occult part of the disease, and there is no simple way to estimate those parameters. It is known [35,89,98,99] that the unavailability of data on the occult stages is due to the fact that almost all detected individuals go under treatment as soon as they

are screened and found to be in some stages of the disease. Then it remains to be clarified what are those statistics presented in the literature, say the percentage of the people in each age and stage group. A close look at those data reveals that they are on the detected stages of the disease, because there is no way to gather data unless a diseased individual is detected. For instance, assume there are data indicating that a% of people who eventually get the disease are found in state i at age t. The only way to interpret these data is that in the long run, assuming no trend of the disease in time, the steady state probability of being in state i' at time t is a.

If these data are available for a system in which there has been no scheduled screening policy in the process of data collection, then this information would be very useful in estimation of the parameters of the model, i.e., a'_{ijt}, b'_{it} and U'_{it}. To do this, it is necessary to hypothesize a model structure for the occult part of the disease and then check the model's output with data. The model which gives the closest match between output and data would be the one which has more chance of representing the actual phenomenon of the disease in the population.*

The following approach is employed in estimation of the model's parameters. It is decided to use the data as the probabilities of being in the detected states (call them $P[X_t=i']$)** and compare them with

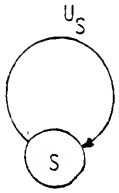
---

*In the use of this distribution, it is assumed that irrespective of the screening time and interval, the population probability of natural detection remains unchanged.

**In this development the prime denoting the probability of natural detection has been dropped. It will be reintroduced later.

the corresponding theoretical values. Due to the dependence of the process on the initial state occupied and time dependency of the disease, there is no long-run distribution. Therefore another concept will be employed in which $\lim_{n\to\infty} P^n_{ijt}$ gives the steady state probability of being in state j after t time intervals, given that the individual started life in state i at time 0; P^n_{ijt} could be computed from postmultiplication of matrix P^{n-1} by matrix P.

Assume that there is no more than one transition per unit interval. This is possible by making the interval small enough to cover the possibility of any short duration transition. Let $P[X_n = i']$ denote the probability of an individual of age n being detected in state i'. The following observations are used to establish a mathematical form for $P[X_t = i']$.



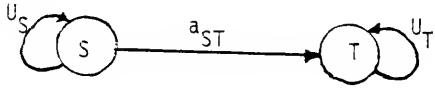
Observation 1:

$$P[X_n=S|X_o=S] = \prod_{j=1}^{n} U_{sj} \qquad (I)$$

Proof: The only possible way for an individual to stay in the state he was in n periods ago, is to stay there at each and every interval, which means the product of the probability of being there at the end of his first, second, ..., and n-th age period.

Observation 2: If there is only one step in going from S to T (i.e., one arc), then

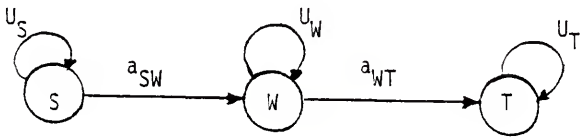
$$P[X_n=T|X_o=S] = \sum_{j=0}^{n-1} \left\{ \left(\prod_{i=1}^{j} U_{si} \right) a_{ST,j+1} \left(\prod_{k=j+2}^{n} U_{Tk} \right) \right\} \qquad (II)$$



Proof: If the individual wants to be at T by age n, he has to jump from S to T sometime during (0,n). Thus he has an extra time of n-1 periods to spend in any possible combination of states S and T.

Observation 3: If there are two steps in going from S to T (i.e., two arcs), then

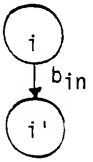
$$P[X_n=T|X_o=S] = \sum_{j=0}^{n-2} \sum_{\ell=j}^{n-2} \left(\prod_{i=1}^{j} U_{Si} \right) a_{SW,j+1} \left(\prod_{k=j+2}^{\ell+1} U_{Wk} \right) \cdot a_{WT,\ell+2} \left(\prod_{m=\ell+3}^{n} U_{Tm} \right) \qquad \text{(III)}$$



Proof: If the individual is to be at T by age n, he has to jump from S to W and from W to T sometime during (0,n). The rest of the time (n-2 remaining periods) has to be spent by staying in any possible combination of states S, W and T. The same type of discussion could be used to generalize the observation for cases of higher order.

Observation 4:

$$P[X_n=i'] = P[X_{n-1}=i] \cdot b_{in} \qquad \text{(IV)}$$



Proof: Due to the assumption employed before (there is no mistake in stage recognition in the detection techniques), there is only one way to be detected in stage i at time n; the individual should have been in state i at time n-1.

The same equality could be derived by using the definition of conditional probability

$$P[X_n=T] = \sum_{\text{all possible S's}} P[X_n=T|X_{t_o}=S] \cdot P[X_{t_o}=S] \qquad 0 \leq t_o \leq n \qquad (1)$$

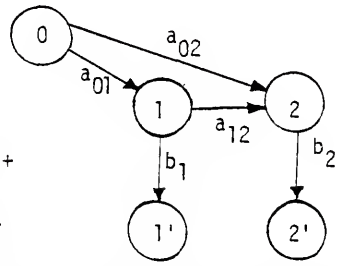
because for the special case when $t_o = n-1$

$$P[X_n=i'\ ] = \sum_{\mu} P[X_n=i'|X_{n-1}=u]\ P[X_{n-1}=\mu] = b_{in} \cdot P[X_{n-1}=i]$$

This analysis shows that for $P[X_t=i']$, the $\lim_{n\to\infty} P^n$ converges to the limit for n=t. The reason behind this is that nobody who is in any occult stage after t-1 can be detected at t, and if somebody has been detected before time t, he would not account for $P[X_t = i']$.

Use of Observations 1 through 4 plus a special case of equation 1 (where $t_o = 0$) makes it possible to find a mathematical form for $P[X_n=i]$ and $P[X_n=i']$ as a function of U's, b's and a's.

As an example, consider a disease which has been postulated to have the following structure:



$$P[X_n=1'] = b_{1n} \cdot P[X_{n-1}=1]$$

$$= b_{1n} \cdot \Big\{ P[X_{n-1}=1|X_o=0]\ P[X_o=0] + P[X_{n-1}=1|X_o=1]\ P[X_o=1] \Big\}$$

where $P\{X_{n-1}=1|X_o=0\}$ and $P\{X_{n-1}=1|X_o=1\}$ could be found from Observations 2 and 1 respectively.

$$P[X_n=2'] = b_{2n} \cdot P[X_{n-1}=2]$$

$$= b_{2n} \cdot \Big\{ P[X_{n-1}=2|X_o=0]\ P[X_o=0] + P[X_{n-1}=2|X_o=1]\ P[X_o=1] + P[X_{n-1}=2|X_o=2]\ P[X_o=2] \Big\}$$

where $P[X_{n-1}=2|X_o=0]$ is the sum of two different conditional probabilities

a) going from 0 to 2 straight (one arc)

b) going from 0 to 2 through 1 (two arcs)

Observations 2 and 3 give expressions for these probabilities respectively. Moreover $P[X_{n-1}=1'|X_o=1]$ and $P[X_{n-1}=2'|X_o=2]$ are found from Observations 2 and 1. Therefore having initial conditions $P[X_o=0]$, $P[X_o=1]$ and $P[X_o=2]$, it would be easy to find the steady state probabilities as functions of the elements of the probability transition matrix.

The question is how can $P[X_n=i']$ be used to estimate the model parameters? A close look at the structure of the transition probability matrix reveals that depending on how the occult part of the disease is to be modeled, a different number of parameters are to be estimated. Assume that the number of parameters of the model of interest is M. Using the methodology developed, the following algorithm will give the unknown parameters uniquely or as functions of other parameters whose estimates would be necessary. The following three concepts are employed in the establishment of the algorithm

1) For each time interval (say n) there are N equations (one for each occult stage) of the form $P[X_n=I'] = \alpha_{I_n}$, where α_{I_n} is the value obtained from data and $P[X_n=I']$ is the mathematical function developed in this section.

2) For each time interval (say n) there are N+1 equations of the form $\sum_j P_{ijn} = 1$; one for state "healthy" and N for N occult stages.

3) For each time interval (say n) there are M unknowns of the form a_{ijn}, b_{in} and U_{in}.

Therefore there are a total of 2N+1 equations and ($M \geq 2N+1$) unknowns.

Algorithm

Step 1: Start with t=1.

Step 2: a) There are 2N+1 equations and m unknowns. Depending on the number of arcs (transitions possible) there may or may not be a unique solution. If not M-(2N+1) of the parameters must be realistically estimated.*

b) Solve the system of equations for the remaining unknowns.

Step 3: Let t = t+1, go back to step 2, unless t = T_{max}, in which case stop. Therefore starting with first period and going in steps of one period, the algorithm gives systematically the value of parameters uniquely or as functions of M-(2N+1) other parameters. Therefore, the algorithm solves the system of equations and determines the remaining unknown parameters, uniquely.

3.5 Objective Functions

Having structured an analytic model of the disease, policy selection can be approached in a straightforward fashion. It should be noted that there are several possible measures of performance; different measures may lead to different "optimal" screening policies. The most common objectives used in the policy selection of cancer screening are

*One way to get an estimate of these parameters is to find a bound on them by estimating the corresponding parameter in the detected stage. This method will be discussed later in this research.

1) Minimization of the total cost--Cost of screening, cost of treatment and cost of losing a patient due to death.

2) Maximization of life expectancy.

3) Maximization of the probability of detection of the disease in a favorable stage.

4) Maximization of the lead time--Time between the detection of the disease under screening and the time that disease would have been discovered under no screening program.

Cost-effectiveness measures of screening policies are an objective of much interest, because if screening is to be used nationwide, there should be a net positive benefit (measured in dollars) of doing so. The total cost of a screening program consists of the cost of screening, the cost of treatment, the cost attributed to the period of disability of a patient and the cost to the society of losing an individual. Therefore

Expected total cost of a screening program = $E(S)+E(T)+E(D)$

where $E(S)$ = Expected total cost of screening

$E(T)$ = Expected total cost of treatment

$E(D)$ = Expected total cost of death due to disease.

In this section, each of these cost elements is expressed mathematically. In order to do this, an undiscounted analysis is employed and later on a discounted version is used to transform all the costs to present worth.

a) Screening Cost

Assume that M screenings are scheduled for ages $T_1, T_2, \ldots, T_M$. The expected total screening cost is the sum of the expected costs of screening at $T_1, T_2, \ldots,$ and T_M. However, since an individual will be screened at age t only if he is in an occult stage (including state "healthy"), the following expected value analysis is used.

Expected cost of screening at age t =

(cost of screening at age t)(probability of being in an occult stage at age t)(probability of test done at age t)

Let $C_s(t)$ = screening cost at age t

$P[X_t=i]$ = probability of being in an occult stage i at age t.

Then

$$\text{Expected cost of screening at age } t = \sum_{i\epsilon 0} C_s(t) \cdot P[X_t=i] \cdot Z_t$$

where 0 is the set of occult stages, i.e., $0 = \{0,1,2,\ldots,N\}$.

The expected total screening cost is the sum of this quantity over all intervals, i.e.,

$$E(S) = \sum_{t=0}^{T_{max}} \left\{ \sum_{i\epsilon 0} C_s(t) \cdot P[X_t=i] \cdot Z_t \right\}$$

b) Treatment Cost

Treatment cost will occur whenever the disease is diagnosed on the screening tests or clinically surfaced. At a specific interval (t,t+1) the expected treatment cost E(t) is

E(t) = (cost of treatment of an individual in stage i' at age t)
(probability of being in a detected stage i' at age t)
(probability of accepting the treatment)

Assume that every individual when detected accepts treatment, and define

$C_T(t,i')$ = Treatment cost of an individual who is in stage i' at age t.

$P[X_t=i']$ = Probability of being in a detected stage i' at age t.

Then

$$E(t) = \sum_{i' \in D} C_T(t,i') \cdot P[X_t=i']$$

where D is the set of detected stages, i.e., $D \in \{1',2',\ldots,N'\}$.

The expected total treatment cost is the sum of this quantity over all intervals, i.e.,

$$E(T) = \sum_{t=0}^{T_{max}} \left\{ \sum_{i' \in D} C_T(t,i') \cdot P[X_t=i'] \right\}$$

c) Mortality Cost

In the simplified model, there is no possibility of death due to disease before detection, and any individual who is in a detected state i' at age t may die with probability $d_{i'}(t)$ and/or be cured with probability $[1 - d_{i'}(t)]$. This transition is assumed to happen one interval after he gets into the detected stage. (This is not what happens in reality, but it is a good approximation of the actual process and is employed here because most of the data on survival are given by age at diagnosis.)

Expected cost of death due to disease at age t = (cost of death at age t)(probability of being in a detected stage i' at age t)(probability of death at age t in stage i')

Let $C_d(t)$ = Cost of death at age t (present worth of future income)

$d_{i'}(t)$ = Probability of death due to disease for an individual who is in stage i' at age t.

Expected cost of death due to disease at age t =

$$\sum_{i' \in D} C_d(t) \cdot P[X_t=i'] \cdot d_{i'}(t)$$

The expected total mortality cost due to disease is the sum of this quantity over all intervals, i.e.,

$$E(D) = \sum_{t=0}^{T_{max}} \left\{ \sum_{i' \varepsilon D} C_d(t) \cdot P[X_t=i'] \cdot d_{i',}(t) \right\}$$

Since this model is based on the assumption that the individual does not die due to any other causes up to age t, let $P_L(t)$ represent the probability of this event, then $P_L(t)$ has to be incorporated in the objective function. Therefore

Expected discounted total cost =

$$\sum_{t=0}^{t_{max}} \left\{ \sum_{i \varepsilon 0} C_s(t) \cdot P[X_t=i] \cdot Z_t + \sum_{i' \varepsilon D} [C_T(t,i') + C_d(t) \cdot d_{i',}(t)] \cdot P[X_t=i'] \right\} \frac{P_L(t)}{(r+1)^t}$$

where r = discount rate

$P_L(t)$ = probability of not dying of other causes up to age t.

It is shown in Section 3.4 that $P[X_t=i]$ and $P[X_t=i']$ are known in terms of the parameters of the model and the decision variables Z_t. It remains to give a realistic estimate of quantities $C_s(t)$, $C_T(t,i')$, $C_d(t)$, $d_{i',}(t)$ and $P_L(t)$.

3.6 An Alternative Expression for the Objective Function

The criterion of minimization of the expected total cost resulted in an objective function derived in the previous section and repeated below.

$$O.F. = \sum_{t=0}^{t_{max}} \left\{ \sum_{i\varepsilon 0} C_s(t)\cdot P[X_t=i] \cdot Z_t + \right.$$

$$\left. + \sum_{i\varepsilon D} [C_T(t,i')+C_d(t)\cdot d_i'(t)] \cdot P[X_t=i'] \right\} \frac{P_L(t)}{(r+1)^t}$$

where $C_s(t)$, $C_T(t,i')$, $C_d(t)$, $d_i'(t)$, $P_L(t)$ and r are the known parameters of the optimization problem and $P[X_t=i]$ and $P[X_t=i']$ are functions of the decision variables Z_t - $0 \leq t \leq t_{max}$.

For any site-specific disease, the methodology developed in Section 3.4 could be employed to write $P[X_t=i]_{i\varepsilon 0}$ and $P[X_t=i']_{i'\varepsilon D}$ as functions of the disease parameters a_{ij}, U_i, b_i and the decision variables Z_t. To do this in the formulation of $P[X_t=i]$ and $P[X_t=i']$, the parameters are written as functions of Z_t in the following manner:

$$a_{ijt} = a'_{ijt} \cdot [1-Z_t\cdot f_i(t)] \qquad i\neq 0$$

$$U_{it} = U'_{it} \cdot [1-Z_t\cdot f_i(t)] \qquad i\neq 0$$

$$b_{it} = Z_t\cdot f_i(t) + [1-Z_t\cdot f_i(t)]\cdot b'_{it}$$

where a'_{ijt}, U'_{it} and b'_{it} stand for the probability elements estimated for the case of no screening program available and a_{ijt}, U_{it} and b_{it} are the corresponding values for any specific screening policy. In other words, prime indicates measurement of elements of the disease when the nature of the disease is unchanged by having no external screening test.

This reduces the objective function to a function of Z_t, $0 \leq t \leq t_{max}$, where Z_t can take one of the two integer values zero or one. Therefore the optimization problem is an integer program but because of the multiplication of variables -- to see this look at any one of the

expressions for $P[X_t=i]$ when a_{ij}, U_i and b_i's are functions of Z_t--it is not an integer linear program and the classic methods of solution to "integer programming" are not applicable. It remains to evaluate all possible policies or use some branch and bound procedure to omit any policy as soon as it is dominated by another policy already computed.

The optimization problem is whether or not to screen the individual at any time t. Therefore if t is a discrete time interval and varies from 1 to t_{max}, there would be $(2)^{t_{max}}$ solutions, because at any time t either there should be a test or should not; of which the best one is to be chosen.

There may be constraints on the total number of tests permitted per each individual and on the interval between two successive tests due to the fact that certain tests are dangerous in nature and/or might introduce some side effects. The constraints are mathematically shown as

Constraint 1: Maximum number of tests permitted for an individual is S_{max}.

$$\sum_{t=1}^{t_{max}} Z_t \leq S_{max}$$

Constraint 2: The interval between two screens should be at least t_{min}.

$$\sum_{i=j}^{j+t_{min}} Z_i \leq 1 \qquad j = 1,2,\ldots,t_{max}-t_{min}$$

Depending on the values of S_{max} and t_{min}, Constraint 2 may imply Constraint 1. For instance if $t_{max} = 72$, $S_{max} = 10$, $t_{min} = 9$, then there is no way to do ten tests or more and keep the interval between two successive tests as desired, which means in this case Constraint 1 is implicitly satisfied by the second set of constraints.

3.6.1 Alternative Form 1

If it is assumed that $f_i(t)$ which is the probability that a person who has been in state i at age t is properly classified as diseased, is a constant independent of age and stage, then it is possible to simplify the objective function and reduce the computational aspects of the problem significantly. The assumption of constant true positive rate of detection is valid in the case of some of the diseases such as Neuroblastoma and, in general, the range of variation of $f_i(t)$ is sufficiently small that a constant could be used to represent the true positive rate of detection of the test. Later on the constant could be varied over its possible domain and a sensitivity analysis could be used to determine the validity of the assumption.

The assumption $f_i(t) = F$ is basic to the following analysis and is used to write

$$a_{ijt} = a'_{ijt}(1-Z_t \cdot F) \qquad i \neq 0$$

$$U_{it} = U'_{it}(1-Z_t \cdot F) \qquad i \neq 0$$

$$b_{it} = b'_{it}(1-Z_t \cdot F) + Z_t \cdot F$$

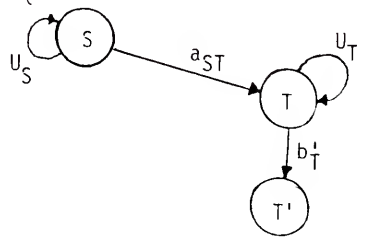
A close look at the mathematical form of $P[X_n=i]$ and $P[X_n=i']$ derived in Section 3.4 shows that each of these terms is the sum of several terms, each of which is in the form of multiplication of n elements. The terms

are such that there exists exactly one element corresponding to each time interval t; $1 \leq t \leq n$. To explain this in more detail, take the following simple case:

Observation 2, Section 3.3 gives

$$P[X_{n-1}=T \mid X_o=S] = \sum_{j=0}^{n-2} \left\{ \left(\prod_{i=1}^{j} U_{Si} \right) a_{ST,j+1} \left(\prod_{k=j+2}^{n-1} U_{Tk} \right) \right\}$$

and Observation 1 gives



$$P[X_{n-1}=T \mid X_o=T] = \prod_{j=1}^{n-1} U_{Tj}$$

Therefore

$$P[X_n=T'] = \left\{ \sum_{j=0}^{n-2} \left\{ \left(\prod_{i=1}^{j} U_{Si} \right) a_{ST,j+1} \left(\prod_{k=j+2}^{n-1} U_{Tk} \right) \right\} P[X_o=S] + \left(\prod_{j=1}^{n-1} U_{Tj} \right) P[X_o=T] \right\} \cdot b_{T,n}$$

$$P[X_n=T'] = \left[\left(a_{ST,1} \cdot U_{T,2} \cdots U_{T,n-1} + U_{S,1} \cdot a_{ST,2} \cdot U_{T,3} \cdots U_{T,n-1} + \cdots U_{S,1} \cdots U_{S,n-2} \cdot a_{ST,n-1} \right) P[X_o=S] + (U_{T1} \cdot U_{T2} \cdots U_{T,n-1}) P[X_o=T] \right] b_{T,n}$$

In this particular example $P[X_n=T']$ is the sum of n+1 terms, each of which consists of multiplications of n elements of the form a_{ij}, U_i and b_i. In each term, say $a_{ST,1} \cdot U_{T,2} \cdot U_{T,3} \cdots U_{T,n-1} \cdot b_{T,n}$, there exists one and only one element corresponding to any time interval $1 \leq t \leq n$ (for t-1, $a_{ST,1}$ is the only parameter that carries the time subscript "1"). This concept plus the assumption of constant $f_i(t)$ will be employed to derive a simpler form for the objective function.

Assume that the screening program is scheduled for $T \varepsilon \{T_1 < T_2 < \ldots < T_M\}$, which means the individual will be screened at ages $T_1, T_2, \ldots,$ and T_M, and use the following notation

$P'[X_t=i]$ = Probability of being in the occult stage i at age t under no screening program.

$P[X_t=i]$ = Probability of being in the occult stage i at age t under any screening program.

$P'[X_t=i']$ = Probability of being in the detected stage i' at age t under no screening program.

$P[X_t=i']$ = Probability of being in the detected stage i' at age t under any screening program.

Then $P'[X_t=i']$ is known from the data and $P'[X_t=i]$ could be found from data as a function of $b_{i,t+1}$ by the use of Observation 4

$$P'[X_t=i] = \frac{P'[X_{t+1}=i']}{b'_{i,t+1}}$$

For any time interval $0 \leq t < T_1$, there is no screening. Therefore

$$P'[X_t=i] = P[X_t=i] \qquad 0 \leq t < T_1$$

$$P'[X_t=i'] = P[X_t=i'] \qquad 0 \leq t < T_1$$

Temporarily, assume that every individual who will eventually get the disease has the disease by the time of the first screening. Later this analysis is used to modify the model to cover every possible case. At time $t = T_1$

$$P[X_{T_1}=i] = (1-F)\, P'[X_{T_1}=i]$$

This is because the last element of $P[X_{T_1}=i]$ carries time T_1 and is modified by (1-F).

For t in the interval $T_1 \leq t < T_2$, one and only one element in $P[X_t=i]$ corresponds to time T_1 and is therefore modified by (1-F). Therefore the whole expression is modified by (1-F).

At time T_2, under the same assumption (everybody has the disease by the time of the first screening), in each term of quantity $P[X_t=i]$ there is one element corresponding to time T_1 and one element corresponding to time T_2. These two elements are the only ones that are modified by the factor (1-F). Therefore the whole expression is modified by $(1-F)^2$, i.e.,

$$P[X_{T_2}=i] = (1-F)^2 P'[X_{T_2}=i]$$

This assumes that the outcome of each screen is independent of previous results. The same analysis holds through for $T_2<t<T_3$ and could be generalized to establish that under the assumption that everybody who eventually gets the disease has it by time T_1, the screening cost will be

$$\sum_{t\in T} C_s(t) \cdot \sum_{i\in 0} P[X_t=i] = \sum_{t\in T} C_s(t) \cdot P'[X_t=0] +$$

$$\left\{(1-F)C_s(T_1) \sum_i P'[X_{T_1}=i] + (1-F)^2 C_s(T_2) \sum_i P'[X_{T_2}=i] + \ldots + \right.$$

$$\left. (1-F)^M C_s(T_M) \sum_i P'[X_{T_M}=i]\right\}$$

This is because the $P[X_t=0]$ is not effected by the screening program; (it is written separately).

Observation 4 can be employed to replace $P'[X_{T_j}=i]$ by its equivalent quantity $P'[X_{T_j+1}=i']/b'_{i,T_j+1}$. Therefore

$$\text{Screening cost} = \sum_{t\in T} C_s(t) \cdot P[X_t=0] + \sum_{k=1}^{M} \left\{ C_s(T_k)\cdot(1-F)^k \cdot \sum_{i'} \frac{P'[X_{T_k+1}=i']}{b'_{i,T_k+1}} \right\}$$

where M is the total number of tests.

To show the relationship between $P[X_t=i']$ and $P'[X_t=i']$ it is necessary to investigate the changes due to b_{it} in addition to those of a_i and U_i and to consider the effect of screening on each individual element of $P[X_t=i']$.

Taking into account only those individuals who get the disease before time T_1 it is obvious that for $0\leq t<T_1$ there is no difference between $P[X_t=i']$ and $P'[X_t=i']$--simply because there is no test in this interval and therefore a_{ij}, U_i and b_i's remain unchanged.

At time $t = T_1$, according to Observation 4

$$P[X_{T_1}=i'] = P[X_{T_1-1}=i] \cdot b_{i,T_1}$$

But from the previous analysis $P[X_{T_1-1}=i] = P'[X_{T_1-1}=i]$, because there has been no test up to T_1. Moreover, since a screening is done at time T_1

$$b_{i,T_1} = b'_{i,T_1} \cdot (1-F) + F$$

Substitution for $P[X_{T_1-1}=i]$ and b_{i,T_1} in $P[X_{T_1}=i']$ gives

$$P[X_{T_1}=i'] = P'[X_{T_1-1}=i] \cdot [b'_{i,T_1}\cdot(1-F)+F]$$

But according to Observation 4

$$P'[X_{T_1-1}=i] = P'[X_T=i']/b'_{i,T_1}$$

$$P[X_{T_1}=i'] = \frac{P'[X_{T_1}=i']}{b'_{i,T_1}} \cdot [b'_{i,T_1}(1-F)+F]$$

$$= P'[X_{T_1}=i'] \cdot [1-F+\frac{F}{b'_{i,T_1}}]$$

For $T_1<t<T_2$ one and only one element in $P[X_t=i']$ corresponds to time $t=T_1$ and this element is either of the set a or U. Therefore the whole expression is modified by factor (1-F), i.e.,

$$P[X_t=i'] = (1-F) \cdot P'[X_t=i'] \qquad T_1<t<T_2$$

At time $t = T_2$, under the assumption that everybody has the disease by the time of the first test, in each term of the quantity $P[X_t=i']$ there is one element corresponding to time T_1 and the term b_{i,T_2} corresponding to time T_2. The element corresponding to time $t=T_1$ is of the set a or U and is therefore modified by factor (1-F) and the element b_{i,T_2} is modified by

$$b_{i,T_2} = b'_{i,T_2} \cdot (1-F) + F$$

Therefore

$$P[X_{T_2}=i'] = (1-F)\, P'[X_{T_2}=i'] \,[1-F+\frac{F}{b'_{i,T_2}}]$$

For $T_2<t<T_3$ there are two elements of the form a and U corresponding to times T_1 and T_2 and each of them is modified by the factor (1-F).

Therefore

$$P[X_t=i'] = (1-F)^2 \cdot P'[X_t=i'] \qquad T_2<t<T_3$$

In a similar manner the whole treatment cost and death cost can be shown to be

$$\text{Treatment cost + death cost} = \sum_{i'\in D} \left\{ \sum_{t=0}^{T_1-1} C_{t,i'} \cdot P'[X_t=i'] + C_{T_1,i'} \left(1-F+\frac{F}{b'_{i,T_1}}\right) \cdot P'[X_{T_1}=i'] + \sum_{t=T_1+1}^{T_2-1} C_{t,i'} \cdot (1-F) \cdot P'[X_t=i'] + C_{T_2,i'} \cdot (1-F)\left(1-F+\frac{F}{b'_{i,T_2}}\right) P'[X_{T_2}=i'] + \ldots \right\}$$

where $C_{t,i'} = C_T(t,i') + C_d(t) \cdot d_{i'}(t)$.

Therefore under the main assumption that all of those individuals who will eventually get the disease are involved at or before time T_1, the expected total cost will be

$$(\text{cost})_1 = \sum_{t=1}^{t_{max}} \left\{ Z_t \cdot C_s(t) \left[P[X_t=0] + (1-F)^{\sum_{k=1}^{t} Z_k} \cdot \sum_{i'} \frac{P'[X_{t+1}=i']}{b'_{i,t+1}} \right] \right\} + \sum_{t=1}^{t_{max}} \left\{ \sum_{i'\in D} \left[C_{t,i'} \cdot (1-F)^{\sum_{k=1}^{t-1} Z_k} \cdot \left(1-F+\frac{F}{b'_{i,t}}\right)^{Z_t} \cdot P'[X_t=i'] \right] \right\}$$

Not everybody has the disease at or before T_1 and, in fact, only a known percentage of them may have the described situation. To find

this percentage, assume that onset time (the time that the individual gets the disease), has a negative exponential distribution with rate λ_o. Then

$$P[X_t \neq 0] = \lambda_o e^{-\lambda_o t}$$

and

$$\int_0^T P[X_t \neq 0]dt = 1 - e^{-\lambda_o \cdot T}$$

Therefore $(1-e^{-\lambda_o \cdot T_1})$ percent of all the individuals who eventually get the disease have had it at or before T_1, and their contribution to objective function is

$$\left(1 - e^{-\lambda_o \cdot T_1}\right) \cdot (\text{cost})_1$$

Using this argument, let us divide the population into M+1 groups--M is the total number of test scheduled--in the following manner:

Group 1 consists of those individuals who have the disease at or before T_1.

Group m consists of those individuals who get the disease between T_{m-1} and T_m. m=2,...,M

Group M+1 consists of those individuals who get the disease after last screening.

From assumption of negative exponential distribution of time of onset, it is clear that

$$P[T_i \leq t \leq T_{i+1}] = e^{-\lambda_o \cdot T_i} - e^{-\lambda_o \cdot T_{i+1}}$$

Considering the second group it is clear that for any individual in this group there would be no benefits from the first screening, but from then on they get the benefit of any other screening. Therefore, using the same analysis as that of group one, it is clear that their contribution to the objective function is

$$\left(e^{-\lambda_o \cdot T_1} - e^{-\lambda_o \cdot T_2}\right) \sum_{t=T_1+1}^{t_{max}} (1-F)^{\sum_{k=T_1+1}^{t} Z_k} \left[\sum_{i' \in D} Z_t \cdot C_s(t) \cdot \frac{P'[X_{t+1}=i']}{b'_{i,t+1}} \right] + \sum_{t=T_1+1}^{t_{max}} \left[\sum_{i' \in D} C_{t,i'} \cdot (1-F)^{\sum_{k=T_1+1}^{t-1} Z_k} \cdot \left(1-F+\frac{F}{b'_{i,t}}\right)^{Z_t} \cdot P'[X_t=i'] \right] \Bigg\}$$

A similar analysis could be used for other groups.

The expected total cost is the sum of all contributions from different groups:

$$O.F. = \sum_{T_j \in T} \left(e^{-\lambda_o \cdot T_{j-1}} - e^{-\lambda_o \cdot T_j}\right) \left\{ \sum_{t=T_{j-1}+1}^{t_{max}} (1-F)^{\sum_{k=T_{j-1}+1}^{t} Z_k} \cdot \left[\sum_{i'} Z_t \cdot C_s(t) \cdot \frac{P'[X_{t+1}=i']}{b'_{i,t+1}} \right] + (1-F)^{\sum_{k=T_{j-1}+1}^{t-1} Z_k} \cdot \left(1-F+\frac{F}{b'_{i,t}}\right)^{Z_t} \cdot C_{t,i} \cdot P'[X_t=i'] \right\} + \sum_{t=1}^{t_{max}} Z_t \cdot C_s(t) \; P[X_t=0]$$

where $T = \{T_k;\ k = 0,\ldots,M+1\}$ and $T_0 = 0$, $T_{M+1} = t_{max}$.

In this formulation Z_t is the decision variable which takes on zero or one, F is the true positive rate of the test, λ_o is the rate of onset of the disease and $b_{i,t}$ is the conditional probability of transition from state i to i' during time (t-1,t). In the objective function λ_o, and therefore $P[X_t=0]$, and $b_{i,t}$ are the only parameters that are not explicitly known from data but the method of Section 3.4 could be used to estimate them.

If, instead of negative exponential, the onset time has a general distribution function $F_{T_o}(t)$, the objective function would be

$$O.F. = \sum_{T_j \in T} \left(F_{T_o}(T_j)-F_{T_o}(T_{j-1})\right) \left\{ \sum_{\tau=T_{j-1}+1}^{t_{max}} (1-F)^{\sum_{k=T_{j-1}+1}^{t} Z_k} \cdot \left[\sum_{i'} Z_t \cdot C_s(t) \cdot \frac{P'[X_{t+1}=i']}{b'_{i,t+1}} \right] + (1-F)^{\sum_{k=T_{j-1}+1}^{t-1} Z_k} \cdot \left(1-F+\frac{F}{b'_{i,t}}\right)^{Z_t} \cdot C_{t,i} \cdot P'[X_t=i'] \right\} + \sum_{t=1}^{t_{max}} Z_t \cdot C_s(t) \cdot P[X_t=0]$$

3.6.2 Alternative Form 2

This method is also used to determine the value of $P[X_t=i]$ and $P[X_t=i']$ as a function of Z_t. The basic idea is the same as the first method and the population is grouped according to the time of onset of the disease. Therefore at each screening time different categories are considered each of which gets a different benefit from screening tests. For instance, for time $t \leq T_1$ there is only one group which gets a

benefit from screening and the cost associated with them at any time t is

$$\sum_i \left\{ C_s(t) \cdot Z_t \cdot (1-F)^{\sum_{k=1}^{t} Z_k} P'[X_t=i] \right\} +$$

$$\sum_i \left\{ C_{t,i'} \left[1-F+ \frac{F}{b'_{i,t}} \right]^{Z_t} \cdot P'[X_t=i'] \right\} \qquad t \leq T_1$$

For any time t ($T_1<t\leq T_2$), there are two groups of individuals

Group 1: Those who had the disease at screening time T_1.

Group 2: Those who had not had the disease at screening time T_1.

The cost associated with the first group at any time t is

$$\sum_i C_s(t) \cdot Z_t \cdot (1-F)^{\sum_{k=1}^{t} Z_k} \left[\sum_j P'[X_t=i, X_{T_1}=j] \right] + \sum_{i'} C_{t,i} (1-F) \cdot$$

$$\left(1-F+ \frac{F}{b'_{i,t}} \right)^{Z_t} \left[\sum_j P'[X_t=i', X_{T_1}=j] \right] \qquad T_1<t\leq T_2$$

where j is the set of occult stages.

The cost associated with the second group at any time t is

$$\sum_i C_s(t) \cdot Z_t \cdot (1-F)^{\sum_{k=T_1+1}^{t} Z_k} \left[\sum_j P'[X_t=i, X_{T_1}=0] \right] + \sum_{i'} C_{t,i'} \cdot$$

$$\left(1-F \frac{F}{b'_{i,t}} \right)^{Z_t} \left[\sum_j P'[X_t=i, X_{T_1}=0] \right] \qquad T_1<t\leq T_2$$

In a similar manner for any time t($T_2<t\leq T_3$) there are three groups of people:

Group 1: Those who were carrying the disease at screening time T_1.

Group 2: Those who were healthy at T_1 but carried the disease by T_2.

Group 3: Those who had not had the disease at time T_2.

The cost associated with these three groups at any time t is

$$\text{a)} \quad \sum_i C_s(t) \cdot Z_t \left\{ (1-F)^{\sum_{k=1}^{t} Z_k} \cdot \left[\sum_j P'[X_t=i, X_{T_1}=j] \right] \right\} + \sum_{i'} \left\{ C_{t,i'} \cdot (1-F)^2 \cdot \left(1-F+\frac{F}{b'_{i,t}} \right)^{Z_t} \cdot \left[\sum_j P[X_t=i, X_{T_1}=j] \right] \right\}$$

$$\text{b)} \quad \sum_i C_s(t) \cdot Z_t \left\{ (1-F)^{\sum_{k=T_1+1}^{t} Z_k} \cdot \left[\sum_j P'[X_t=i, X_{T_1}=0, X_{T_2}=j] \right] \right. + \sum_{i'} \left\{ C_{t,i'} \cdot (1-F) \cdot \left(1-F+\frac{F}{b'_{i,t}} \right)^{Z_t} \cdot \left. \left[\sum_j P'[X_t=i', X_{T_1}=0, X_{T_2}=j] \right] \right\}$$

$$\text{c)} \quad \sum_i C_s(t) \cdot Z_t \left\{ (1-F)^{\sum_{k=T_2+1}^{t} Z_k} \left[P'[X_t=i, X_{T_2}=0] \right] \right\} + \sum_{i'} \left\{ C_{t,i'} \cdot \left(1-F+\frac{F}{b'_{i,t}} \right)^{Z_t} \cdot \left[P'[X_t=i', X_{T_2}=0 \right] \right\}$$

for $T_2 < t \leq T_3$

The total cost is

$$\sum_i \left\{ \sum_{t=0}^{T_1} \left(Z_t \cdot C_s(t) \cdot P'[X_t=i] + (1-F+\frac{F}{b'_{i,t}})^{Z_t} \cdot C_{t,i'} \cdot P'[X_t=i'] \right) + \right.$$

$$\sum_{t=T_1+1}^{t_{max}} (1-F)^{\sum_{k=1}^{t} Z_k - 1} \left(Z_t \cdot C_s(t) \cdot \left[\sum_j P'[X_t=i, X_{T_1}=j] \right] + \right.$$

$$\left. \left. (1-F+\frac{F}{b'_{i,t}})^{Z_t} \cdot C_{t,i'} \cdot \left[\sum_j P'[X_t=i', X_{T_1}=j] \right] \right) +$$

$$\sum_{m=2}^{M} \sum_{t=T_{m+1}}^{t_{max}} (1-F)^{\sum_{k=1}^{t} Z_k - m} \left(Z_t \cdot C_s(t) \cdot \left[\sum_j P[X_t=i, X_{T_{m-1}}=0, X_{T_m}=j] \right] + \right.$$

$$\left. (1-F+\frac{F}{b'_{i,t}})^{Z_t} \cdot C_{t,i'} \cdot \left[\sum_j P[X_t=i', X_{T_{m-1}}=0, X_{T_m}=j] \right] \right) +$$

$$\sum_{t=T_1+1}^{t_{max}} \left(Z_t \cdot C_s(t) \cdot \left[P[X_t=i, X_{T_{g(t)}}=0] \right] + \right.$$

$$\left. \left. (1-F+\frac{F}{b'_{i,t}})^{Z_t} \cdot C_{t,i'} \cdot \left[P[X_t=i', X_{T_{g(t)}}=0] \right] \right) \right\}$$

where $g(t) = \sum_{k=2}^{t} Z_{k-1}$.

To get the form of the objective function it is necessary to find all those joint probability distributions from the data, but there is a pattern in them which categorizes them into three groups:

a) $\sum_j P'[X_t=i, X_{T_1}=j]$

b) $\sum_j P'[X_t=i, X_{T_{m-1}}=0, X_{T_m}=j]$

c) $P'[X_t=i, X_{T_k}=0]$

If the distribution function for the occult part of the disease is known, it would not be difficult to find the general form of these joint probability distributions, but, due to the complexity of the form of objective functions, this model is not implemented.

CHAPTER FOUR

NEUROBLASTOMA

The analysis developed in Chapter Three is based on the fact that in the case of some diseases such as cancer the age of the individual and the stage of the disease at the time of detection are very important and, generally, the younger the individual and the earlier the detection, the better are the results of treatment.

In order to implement the model to a particular disease it was necessary to gather data which support the idea of age and stage effect. During this investigation process it was noted that Neuroblastoma is a disease which is very much age and stage dependent and has other characteristics — such as its duration which is mostly for children below four years old — which makes it a good choice for modeling from a computational point of view. Therefore it was decided to implement the model for this particular disease. The problem with this choice was the rare incidence of Neuroblastoma which made it almost impossible to work with a reasonable amount of data. In most instances, it was necessary to combine different sources of data to get estimates of the parameters. During the period of data collection, it was found that in the past little attention has been paid to collecting enough information for each case. Even in some of the cases the date of birth and sex are unknown [74]. It is surprising that for such a malignant disease as Neuroblastoma, which fortunately has a very low incidence rate, there has been no interest in recording all of the important aspects of the disease such

as,

a) A biography of the case including age, sex, race, family situation and location.

b) The progression of the disease including the stage and time of transition.

c) Follow up procedure up to death of the case.

Even though this research uses the primary data, the model presented is general in nature and it is hoped that the future investigators will gather data in a standard manner such as the one established in this paper.

Section 4.1 is a review of literature on Neuroblastoma. Section 4.2 develops the equations for the unconditional probability of being in a detected stage at age t. Section 4.3 presents the cost-effectiveness objective function. Section 4.4 is on the subject of computerizing the model by using available estimates for the elements of the objective function. Section 4.5 uses sensitivity analysis to vary the unknown parameters over their entire range and study their effects on the optimal policy chosen, and Section 4.6 presents the results and conclusions of the implemented model.

4.1 Literature on Neuroblastoma

According to Miller [72], cancer is the second most common cause of death in childhood in the United States, and among the fifty-two forms of childhood cancer, Neuroblastoma is ranked "the third," Leukemia and Limphoma being the first two. Neuroblastoma is a tumor of childhood, which develops from neural crest tissue and may arise from any anatomic site along the craniospinal axis. It may be present as a large "silent" intra-abdominal mass [72,73,108], which unfortunately metastasizes

rapidly, so that up to 70% of the patients have metastasis at the time of diagnosis [33,42]. Fortner et al. [33] says that Neuroblastoma is seldom manifested before its extent is massive or metastases are distant from the primary site.

Signs and symptoms produced by Neuroblastoma are largely attributable to compression of other structures by the primary tumor or by metastatic deposits [108]. Usually the initial evidence of the disease is a mass [108]. In about 55% of the cases the tumor arises in the abdomen and 37% of all Neuroblastomas arise from an adrenal gland; an additional 18% arise from nonadrenal intra-abdominal sites [12,18,30,33,42,61,73,81]. Figure 7 shows the site of the primary tumor along with cure rates as given by Williams and Donaldson [108].

Although the etiology of Neuroblastoma is unknown, it is known that it is the most common malignant solid tumor in childhood [12,18,108]. Daregon [18] in a review of 1833 cases of childhood malginancies found 205 examples of Neuroblastoma for an incidence of 11%, whereas Bodian [12] noted 129 examples among 907 children (14%) with malignancies, in England and Wales.

About half of the cases occur in children two years of age and younger, and about three-fourths of all cases occur during the first four years of life [108]. Leonard and coworkers believe that about one-third of Neuroblastoma occurs in the first year of life and 15% yearly thereafter until 6 or 7 years of age [64]. Wilson from England claims that 41% of patients are below two years of age and 91% below ten years [51], which is comparable with the data of Breslow and McCann [13], Stella et al. [94], Gross et al. [42] and De Lorimier et al. [20].

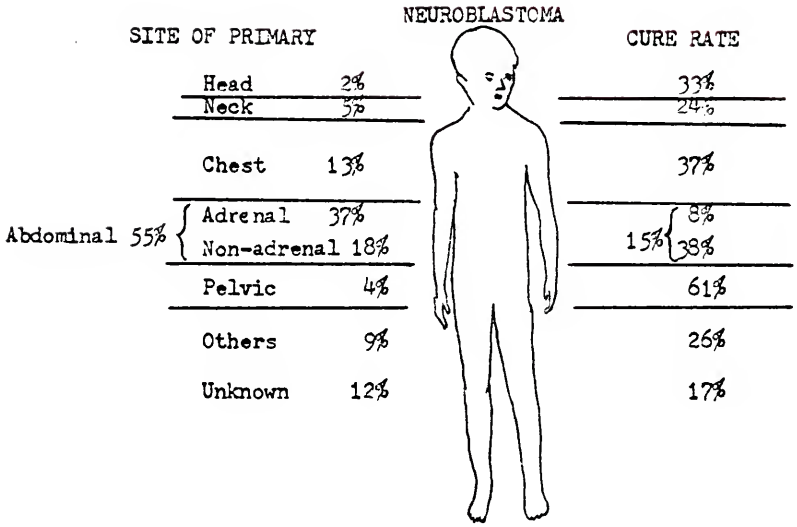


Figure 7: An analysis of composite data from literature.

"By Williams and Donaldson [108]".

The incidence of Neuroblastoma is not yet known and different researchers have different ideas. According to Leonard et al. [64], among a birth cohort of three million children there are 31 Neuroblastoma deaths per year or 155 deaths by the fifth birthday, which is comparable to data from Beckwith and Martin [7] and Peterson et al. [74]. Miller [72] estimated the average annual death rate to be 10 per million for children 0 to 4 years of age and 4 per million for those 5 to 9 years of age. Williams and Donaldson [108] claim that it is possible that true incidence of Neuroblastoma is greater than that reported. Beckwith and Perrin [8] show that based on studies of random sections of the adrenal glands of autopsied infants under three months of age, the incidence of Neuroblastoma is between 1:179 and 1:259. Guin et al. [43] obtained deliberate serial sections of the adrenals and found the incidence to be 1:39, which implies that many infants may have an innate ability to induce spontaneous regression. Everson and Cole have published a book on spontaneous regression of cancer [28], and Neuroblastoma is one of those rare neoplasms in which spontaneous regression is possible [108]. In general, these data show that incidence of Neuroblastoma may be greater than indicated by the statistics, and most likely is in the range of one in 30,000-40,000 infants.

Those researchers, who have been interested in determining factors effective in Neuroblastoma, have searched variation in sex, race, season, year, space-clustering and familial disease. Their conclusion shows that sex is the only effective factor from those mentioned. The male:female ratio of Neuroblastoma death varies from 1.1:1.0 (Miller, [73]), to 1.22:1 (Kinnier Wilson, [51]), which is gathered from data in England and is comparable to 1.29:1 (Makinen, [67]), and 1.26:1 (De Lorimer, [20]).

Even though Lee [63] noted an increased incidence of Neuroblastoma death in England during the summer months, the remaining researchers, including Miller [73] found no seasonal variation in death occurring in the United States. However, Miller noticed a statistically significant year-to-year variation in a few geographic subdivisions, including Texas.

There have been no unequivocal reports of Neuroblastoma in parent and child data [108], and Dinicola and Movassaghi [21] found no significant differences between the white and black children in regard to median duration of survival or percentage of long-term survivors.

It would appear that there has been little improvement in diagnosis and treatment of Neuroblastoma. The subcommittee on Childhood solid tumors, National Cancer Institute (N.C.I.), with Sutow as chairman [96] reported on comparison of survival rates in 1956 versus 1962. They noted that the median age at diagnosis was two years eight months in 1956, two years in 1962, and there were no statistically significant difference in survival curves for children first seen in 1956, as compared to those first seen in 1962. This implies that if the child is detected clinically, he has a little chance of survival. This emphasizes the potential importance of screening.

It has been seen that survival from Neuroblastoma depends on age of the individual and extent of the disease at time of diagnosis. Sutow [96] noted that best survival occurs among those under one year and decreases with increase in age. In general, although a favorable prognosis is seen in children under one to two years of age, according to Williams and Donaldson [108], the overall mortality is high. One reason for high rate of mortality is that disease is metastasized at time of diagnosis. Up to 70% of patients have metastasis at that time (Leonard [64], Sutow [96],

Gross [42]). Sutow [96] found that there is a highly significant difference ($P < 0.005$) in survival rate favoring the patients without metastasis under two years of age, over those older and without metastasis.

Sutow also found that the survival curve for the regional disease group was intermediate between the survival curve for localized and metastatic groups, which shows that extension beyond the organ involved by the primary tumor mass is definitely associated with a less favorable outcome.

Evans [23,24,25,26] showed that in all age categories, survival in children with metastatic disease is significantly inferior to survival among children without metastases.

This general information about the disease helps to model the disease as progressing through different stages. There are at least three different modeling approaches for Neuroblastoma.

In December 1965, the American Joint Committee on Cancer Staging [16] offered the TNM system. This system was based upon the use of the three capital letter: T - Tumor or primary lesion and its extent; N - lymph nodes of the region and their condition; M - distant metastasis. Within each letter element, increasing involvement was categorized by the combination of the capital letter with a numerical suffix. According to Evans [27], in the case of Neuroblastoma TNM seems to be inapplicalbe, because it is well recognized that age is an important prognostic criterion, so that irrespective of the location of the tumor, the younger the child, the better the prognosis (Daregon [18]).

In November 1967, D. H. James [50] published his model, which consists of four different stages,

Stage I, localized - resectable

Stage II, regional - unresectable

Stage III, generalized - without bone or bone marrow involvement

Stage IV, generalized - with bone or bone marrow involvement.

This method of staging has been used by a few investigators, such as Pinkel et al. [76].

In February 1971, A. E. Evans et al. [27] offered a model for Neuroblastoma which, they believed, more reliably described the extent of the disease present. Their model does not require reference to resectability, which itself requires surgical judgment [23,27]. The Evans method is

Stage I: Tumor confined to the organ or structure of origin.

Stage II: Tumors extending in continuity beyond the organ or structure of origin but not crossing the midline. Regional lymph nodes on the homolateral side may be involved.

Stage III: Tumors extending in continuity beyond the midline. Regional lymph nodes may be involved bilaterally.

Stage IV: Remote disease involving skeleton, organs, soft tissues, or distant lymph node groups, etc. (see IV-S).

Stage IV-S: Patients who would otherwise be stage I or II, but who have remote disease confined only to one or more of the following sites: liver, skin, or bone marrow (without radiological evidence of bone metastases on complete skeletal survey).

In proposing this method of staging Evans and D'Angio suggest investigations necessary for the staging of a patient with localized Neuroblastoma,

investigations desirable but not mandatory for staging a patient with Neuroblastoma and additional necessary investigations depending on the tumor size [27].

Significant progress in treatment and screening of Neuroblastoma is due to the discovery of the increased excretion of urinary metabolites of the catecholamines in patients with tumors of the neural crest. The determination of Vanillylmandelic acid (VMA), Homovanillic acid (HVA), Norepinephrine (NE) and Dopamine helps physicians to diagnose Neuroblastoma more easily and gives them an idea in judging the therapeutic effect of surgery, radiotherapy, and chemotherapy [25,44,64,108]. The quantitative measurement of VMA on 24-hour urine collection is the most accurate approach to biochemical diagnosis (Voorhess [102,103,104], Greenberg and Gardner [40], Labrosse [62], Gitlow et al. [37,38]).

The problem with quantitative measurement of VMA is that it is not easy to use it for mass screening and it is difficult to collect a 24-hour urine sample. Therefore the need for a better screening test led Labrosse to develop the VMA spot test [44,64], which could be done with a drop of urine. Evans et al. [25], give two reasons for the superiority of the Labrosse spot test over quantitative VMA tests, which includes simplicity and its good interpretation. They suggest, "When a questionable result is obtained, the patient must be placed on a restricted diet and the test be repeated in 24 hours and again if necessary in 48 hours. If consistently a 'questionable' reaction was shown, a quantitative test should be done" (page 914).

Helson [46] believes that examinations or procedures used in diagnostic screening for Neuroblastoma should be those that are most specific, sensitive and least injurious to the child. No single test fulfills all

these requirements, but Labrosse [62] says that in spite of VMA spot test limitations, the simplicity and speed of this test encourage its application as a screening test on the urine of patients who are suspected of having tumors.

It should be noted that no false sense of security can be derived from negative results since not every patient with Neuroblastoma excretes increased amounts of VMA. This lowers the reliability of the tests. There have been some investigations to measure this factor in the case of Neuroblastoma (Williams and Donaldson [108]). The data in the literature show a reliability factor of 80%. This means that an average of only 80% of those patients having Neuroblastoma will be detected by quantitative VMA.

Leonard and other investigators [64] used the fact that up to 80% of children with Neuroblastoma excrete excess Catecholamine or one of its metabolites, 3-methoxy, 4-hydroxymandelic acide (VMA) in the urine to develop a VMA test strip. The test is based, in part, on the chemical reaction and color change of the spot test to measure its false positive incidence. Leonard used an extensive clinical trial on 20,000 patients. Of 50 patients with positive results, 30 were rejected on the next test. Of the 20 patients remaining, 15 were cancer patients on some type of chemotherapy. The test, therefore, has a false-positive incidence of $\frac{50}{20,000} = \frac{1}{400}$. They suggest that massive screening be carried out in the following manner: at birth and during each physician visit, six times the first year, three times the second year, twice a year thereafter until the child reaches 6 or 7 years of age.

Most of the literature in Neuroblastoma examines the medical characteristics of the disease-- its origin, diagnosis, treatment, and survival

rate [1,7,33,41,42,43,47,48,51,58,60,67,70,76,77,81,96,100,107]. The only statistical investigation in this area is done by Norman Breslow and Barbara McCann [13]. They used variant of a linear logistic model to analyze two years survival proportions for 245 children treated for Neuroblastoma. Their analysis showed that age and stage of the disease at diagnosis are important factors in determination of chance of survival. They used Evans' method of staging [27] and treated age as a continuous variable.

Breslow and McCann used data from two different sources, and therefore had to "smooth" them. The smoothing was based on the logistic transformation of the survival probability, P, associated with any particular age, stage category. The model is

$$\ell n\left(\frac{P}{1-P}\right) = \mu + \alpha + \beta$$

$$\Rightarrow \quad P = [1+\exp(-\mu-\alpha-\beta)]^{-1}$$

where μ is a constant, and α and β are the additive stage and age effects. α and β are assumed to add up to zero over stage and age categories, respectively. They used the method of maximum likelihood to estimate α, β and μ. The results are presented in Table 1.

The equation used for smoothing assumes that the effect of age and stage on survival probabilities is additive in logistic scale, which means there is no interaction between α and β. The smoothed survival proportions for the data are presented in Table 2. This analysis adjusted for the effects of age by means of grouping into three broad categories. Treating age as a continuous variable t, measured in months, they obtained a refined analysis which could result in more accurate

Table 1: Maximum likelihood estimate of the parameters in model

$$\ell n\left(\frac{1}{1-p}\right) = \mu + \alpha + \beta$$

Overall mean	Stage effects	Age effects
$\mu = 0.057 \pm 0.227$	$\alpha_I = 1.684 \pm 0.568$	$\beta_{0-11} = 1.577 \pm 0.316$
	$\alpha_{II} = 0.431 \pm 0.384$	$\beta_{12-23} = -0.541 \pm 0.323$
	$\alpha_{III} = -0.092 \pm 0.414$	$\beta_{24+} = -1.036 \pm 0.280$
	$\alpha_{IV} = -2.684 \pm 0.387$	
	$\alpha_{IVS} = 0.661 \pm 0.472$	

Table 2: Smoothed survival proportions by age and stage of the disease.

Age (months)	Stage				
	I	II	III	IV	IVS
0-11	0.965	0.887	0.824	0.259	0.909
12-23	0.768	0.487	0.360	0.040	0.544
24+	0.669	0.366	0.255	0.025	0.421

prognosis. They tested six different models shown in Table 3. Models A and D gave better results and were therefore preferred.

The probability of survival for two years free of disease estimated on the basis of Models A and D are presented in Figure 8.

The only cost analysis is done by George Hallett [44] in a brief letter to the editor of the Journal of Pediatric Surgery. He uses a rough estimate of the cost per patient using the screening policy mentioned by Leonard et al. [64] and comes up with $200,000 cost of detecting one patient. He asks whether or not this screening policy is to the benefit of society. This brings out the question of what is a better way of doing screening, which is the purpose of this research.

4.2 Development of Unconditional Probabilities

Earlier in this chapter some of the different staging methods used in the process of staging Neuroblastoma were mentioned. In this research Evans' method of staging is used, according to which, there are five different stages of the disease:

Stage I: Tumor confined to the organ or structure of origin.

Stage II: Tumors extending in continuity beyond the organ or structure of origin but not crossing the midline.

Stage III: Tumors extending in continuity beyond the midline.

Stage IV: Remote disease involving skeleton, organs, soft tissues or distant lymph node group, etc.

Stage IV-S: Patients who would otherwise be stage I or II, but who have remote disease confined only to one or more of the following sites: liver, skin, or bone marrow.

Each of these stages could be occult of detected. In addition, there

Table 3: Different models for continuous age. x is age in months at diagnosis while i=I,II,III,IV,IVS

MODEL	Equation for $Ln(\frac{P}{1-P})$	Number of parameters estimated
A	$a_i + b \cdot x$	6
B	$a_i + b \cdot x + c \cdot x^2$	7
C	$a_i + b_i \cdot x$	10
D	$a_i + d \cdot Ln(x+6)$	6
E	$a_i + d \cdot Ln(x+6) + e \cdot Ln^2 (x+6)$	7
F	$a_i + d_i \cdot Ln(x+6)$	10

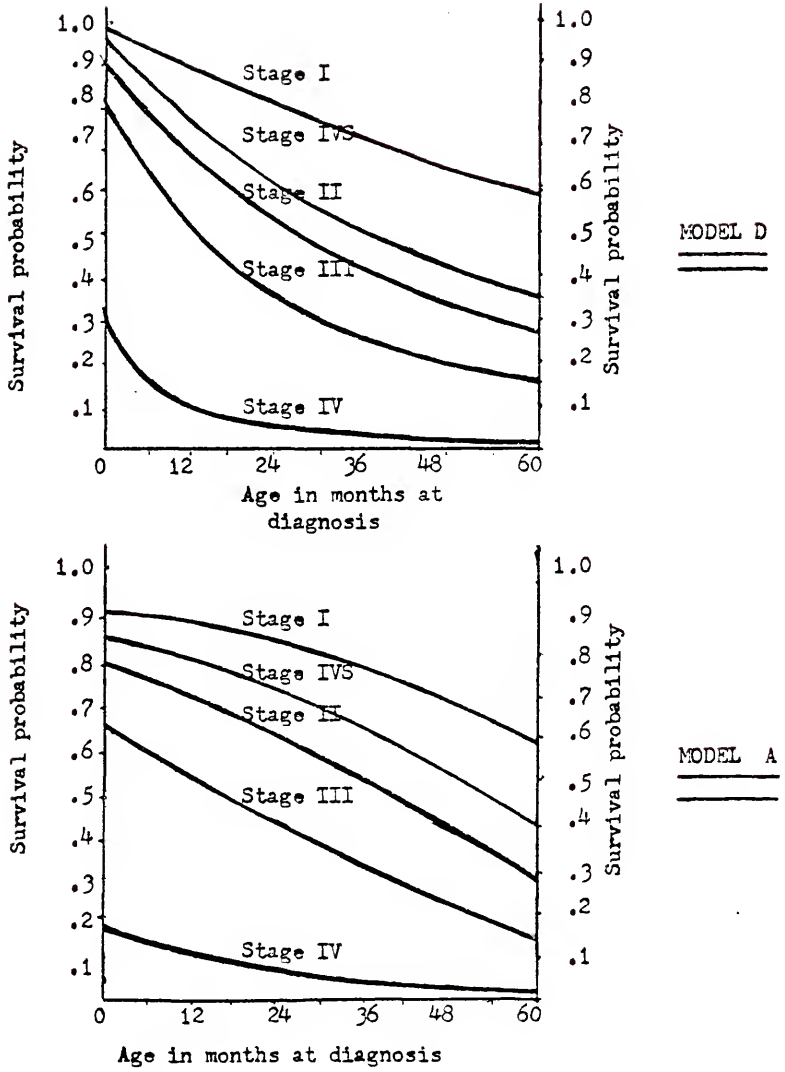


Figure 8: Probability of survival estimated on the basis of model A and D.

are three states corresponding to "free of disease," "cured" and "death." Therefore for any age under consideration there are 13 stages.

The model shown in Figure 9 is postulated to represent the disease progression in an individual. As in the general model, regardless of the actual progression of the disease in the detected stages, the individual is assumed to die from the disease or be "cured."*

In what follows a_{ij}, b_i and U_i are defined as in the general case and P' is used to indicate the probability under consideration in the case of no screening policy. To clarify this notation it should be mentioned that in most of the diseases, including Neuroblastoma, the probability elements of the model are known for the case when there has been no external action such as screening applied to the system. For each time interval t, there are 16 unknowns as follows,

$$U_o, U_I, U_{II}, U_{III}, U_{IV}, U_{IV\text{-}S}$$

$$b_I, b_{II}, b_{III}, b_{IV}, b_{IV\text{-}S}$$

$$a_{01}, a_{12}, a_{23}, a_{34}, a_{35}$$

all of which are in the occult part of the disease and an explicit estimation of which is impossible. Therefore the method developed in Section 3.4 is employed to write $P[X_t = i]$ and $P[X_t = i']$ i = I,II,III, IV,IV-S as a function of a_{ij}, U_i and b_i's. The analysis is as follows. Let

$$P[X_o = O] = \delta_o \ , \ P[X_o = I] = \delta_1, \ \ldots, \ P[X_o = IV\text{-}S] = \delta_5$$

*In the case of Neuroblastoma as well as in the general case, the spontaneous regression of the disease has not been considered--due to the difficulty of obtaining data on this phenomenon.

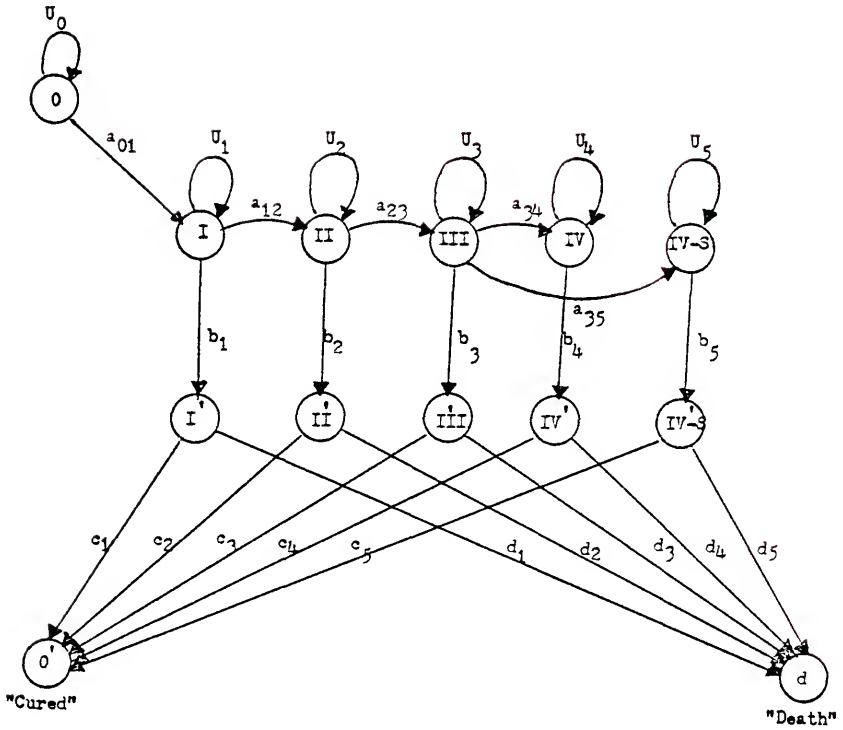


Figure 9: Model of the disease (NEUROBLASTOMA), representing progression of the disease in an individual.

Then,

$$P[X_n=I'] = P[X_n=I'|X_o=0]\ P[X_o=0] + P[X_n=I'|X_o=I]\ P[X_o=I]$$

but,

$$P[X_n=I'|X_o=0] = \left[\sum_{j=0}^{n-2}\left(\prod_{i=1}^{j} U_{oi}\right)\cdot a_{01,j+1}\cdot\left(\prod_{k=j+2}^{n-1} U_{1k}\right)\right] b_{1,n}$$

"observation 2"

and

$$P[X_n=I'|X_o=1] = \left(\prod_{j=1}^{n-1} U_{1j}\right) b_{1,n} \qquad \text{"observation 1"}$$

$$\therefore \quad P[X_n=I'] = \left\{\left[\sum_{j=0}^{n-2}\left(\prod_{i=1}^{j} U_{oi}\right)\cdot a_{01,j+1}\cdot\left(\prod_{k=j+2}^{n-1} U_{1k}\right)\right]\delta_o + \left(\prod_{j=1}^{n-1} U_{1j}\right)\delta_1\right\}\cdot b_{1,n}$$

Also,

$$P[X_n=II'] = P[X_n=II'|X_o=0]\ P[X_o=0] + P[X_n=II'|X_o=I]\ P[X_o=I] + P[X_n=II'|X_o=II]\ P[X_o=II]$$

but $P[X_n=II'|X_o=0]$ is known from Observation 3, $P[X_n=II'|X_o=I]$ is known from Observation 1 and $P[X_n=II'|X_o=II]$ is known from Observation 2. Therefore,

$$P[X_n=II'] = \left\{\left[\sum_{j=0}^{n-3}\sum_{\ell=j}^{n-3}\left(\prod_{i=1}^{j} U_{oi}\right)\cdot a_{01,j+1}\cdot\left(\prod_{k=j+2}^{\ell+1} U_{1k}\right)\cdot a_{12,\ell+2}\cdot\left(\prod_{m=\ell+3}^{n-1} U_{2k}\right)\right]\delta_o +$$

$$\left[\sum_{j=0}^{n-1}\left(\prod_{i=1}^{j} U_{1i}\right)\cdot a_{12,j+1}\cdot\left(\prod_{k=j+2}^{n-1} U_{2k}\right)\right]\delta_1 +$$

$$\left[\left[\prod_{j=1}^{n-1} U_{2j}\right]\delta_2\right\}b_{2,n} \qquad n\geq 2$$

A similar analysis could be used to write $P[X_n=III']$, $P[X_n=IV']$ and $P[X_n=IV\text{-}S']$ as a function of elements a_{ij}, U_i and b_i's. The result is,

$$P[X_n=III'] = \left\{\left[\sum_{j=0}^{n-4}\sum_{\ell=j}^{n-4}\sum_{P=\ell+1}^{n-3}\left(\prod_{i=1}^{j} U_{oi}\right)\cdot a_{o1,j+1}\cdot\right.\right.$$

$$\left.\left(\prod_{k=j+2}^{\ell+1} U_{1k}\right)a_{12,\ell+2}\left(\prod_{m=\ell+3}^{P+1} U_{2m}\right)a_{23,P+2}\left(\prod_{q=P+3}^{n-1} U_{3q}\right)\right]\delta_o +$$

$$\left[\sum_{j=0}^{n-3}\sum_{\ell=j}^{n-3}\left(\prod_{i=1}^{j} U_{1i}\right)a_{12,j+1}\left(\prod_{k=j+2}^{\ell+1}\right)a_{23,\ell+2}\cdot\right.$$

$$\left.\left(\prod_{m=\ell+3}^{n-1} U_{3m}\right)\right]\delta_1 +$$

$$\left[\sum_{j=0}^{n-2}\left(\prod_{i=1}^{j} U_{2i}\right)a_{23,j+1}\left(\prod_{k=j+2}^{n-1} U_{3k}\right)\right]\delta_2 +$$

$$\left.\left(\prod_{j=1}^{n-1} u_{3j}\right)\delta_3\right\}b_{3,n} \qquad n\geq 2$$

$$P[X_n=IV'] = \Bigg\{ \delta_o \Bigg[\sum_{j=0}^{n-5} \sum_{\ell=j}^{n-5} \sum_{p=\ell+1}^{n-4} \sum_{R=P+1}^{n-3} \Big(\prod_{i=1}^{n} U_{oi} \Big) a_{o1,j+1} \cdot$$

$$\Big(\prod_{k=j+2}^{\ell+1} U_{1k} \Big) a_{12,\ell+2} \Big(\prod_{m=\ell+3}^{P+1} U_{2m} \Big) a_{23,P+2} \Big(\prod_{q=P+3}^{R+1} U_{3q} \Big) a_{34,R+2}$$

$$\Big(\prod_{S=R+3}^{n-1} U_{4S} \Big) \Bigg] +$$

$$\Bigg[\sum_{j=0}^{n-4} \sum_{\ell=j}^{n-4} \sum_{p=\ell+1}^{n-3} \Big(\prod_{i=1}^{j} U_{1i} \Big) a_{12,j+1}$$

$$\Big(\prod_{k=j+2}^{\ell+1} U_{2k} \Big) a_{23,\ell+2} \Big(\prod_{m=\ell+3}^{P+1} U_{3m} \Big) a_{34,P+2} \Big(\prod_{q=P+3}^{n-1} U_{4q} \Big) \Bigg] \delta_1 +$$

$$\Bigg[\sum_{j=0}^{n-3} \sum_{\ell=j}^{n-3} \Big(\prod_{i=1}^{j} U_{2i} \Big) a_{23,j+1} \Big(\prod_{k=j+2}^{\ell+1} U_{3k} \Big) a_{34,\ell+2}$$

$$\Big(\prod_{m=\ell+3}^{n-1} U_{4m} \Big) \Bigg] \delta_2 +$$

$$\Bigg[\sum_{j=0}^{n-2} \Big(\prod_{i=1}^{j} U_{3i} \Big) a_{34,j+1} \Big(\prod_{k=j+2}^{n-1} U_{4k} \Big) \Bigg] \delta_3 +$$

$$\Big(\prod_{j=1}^{n-1} U_{4j} \Big) \delta_4 \Bigg\} b_{4,n}$$

and

$$P[X_n=IV\text{-}S'] = \Bigg\{\Bigg[\sum_{j=0}^{n-5}\sum_{\ell=j}^{n-5}\sum_{P=\ell+1}^{n-4}\sum_{R=P+1}^{n-3}\Big(\prod_{i=1}^{k}U_{oi}\Big)a_{01,j+1}\cdot$$

$$\Big(\prod_{k=j+2}^{\ell+1}U_{1k}\Big)a_{12,\ell+2}\Big(\prod_{m=\ell+3}^{P+1}U_{2m}\Big)a_{23,P+2}\Big(\prod_{q=P+3}^{R+1}U_{3q}\Big)a_{35,R+2}\cdot$$

$$\Big(\prod_{S=R+3}^{n-1}U_{5S}\Big)\Bigg]\delta_0 +$$

$$\Bigg[\sum_{j=0}^{n-4}\sum_{\ell=j}^{n-4}\sum_{P=\ell+1}^{n-3}\Big(\prod_{i=1}^{j}U_{1i}\Big)a_{12,j+1}\cdot$$

$$\Big(\prod_{k=j+2}^{P+1}U_{2k}\Big)a_{23,\ell+2}\Big(\prod_{m=\ell+3}^{P+1}U_{3m}\Big)a_{35,P+2}\Big(\prod_{q=P+3}^{n-1}U_{5q}\Big)\Bigg]\delta_1 +$$

$$\Bigg[\sum_{j=0}^{n-3}\sum_{P=j}^{n-3}\Big(\prod_{i=1}^{j}U_{2i}\Big)a_{23,j+1}\Big(\prod_{k=j+2}^{\ell+1}U_{3k}\Big)a_{35,\ell+2}\cdot$$

$$\Big(\prod_{m=\ell+3}^{n-1}U_{5m}\Big)\Bigg]\delta_2 +$$

$$\Bigg[\sum_{j=0}^{n-2}\Big(\prod_{i=1}^{j}U_{3i}\Big)a_{35,j+1}\Big(\prod_{k=j+2}^{n-1}U_{5k}\Big)\Bigg]\delta_3 +$$

$$\Big(\prod_{j=1}^{n-1}U_{5j}\Big)\delta_5\Bigg\}b_{5,n}$$

It is seen that if δ_i is zero for any i, then the form of $P[X_n=i']$ reduces significantly. For the time being, there are not enough data to support the idea that $\delta_i = 0$ for some i. Therefore it is necessary to account for any possible condition at birth.

Using the algorithm presented in Section 3.4, it is possible to go in steps of one and get the value of a_{ij}, b_i and U_i's as functions of $P'[X_t=i']$. This is because for each time interval (t-1,t) there are 11 equations of the form,

$$P'[X_t=i'] = \alpha_i, \qquad i' = \{I', II', III', IV', IV\text{-}S'\}$$

$$U_{ot} + a_{olt} = 1$$

$$U_{1t} + a_{12t} + b_{1t} = 1$$

$$U_{2t} + a_{23t} + b_{2t} = 1$$

$$U_{3t} + a_{34t} + a_{35t} + b_{3t} = 1$$

$$U_{4t} + b_{4t} = 1$$

$$U_{5t} + b_{5t} = 1$$

and therefore there is a system of 11 equations and 16 unknowns. It is seen that unless there is an accurate estimate for at least 5 = 16-11 of the unknowns (say b's) the system of equations has more than one solution and in fact there are infinite sets of parameters that could be chosen to produce the desired $P'[X_t=i']$.

4.3 Objective Function

The objective function is the same as the general case developed in Section 3.5, i.e.,

$$O.F. = \sum_{t=0}^{t_{max}} \left\{ \sum_{i\varepsilon O} C_s(t) \cdot P[X_t=i] \cdot Z_t + \sum_{i'\varepsilon D} [C_T(t,i') + C_d(t) \cdot d_i(t)] \cdot P[X_t=i'] \right\} \frac{P_L(t)}{(r+1)^t}$$

where $O \in \{0, I, II, III, IV, IV\text{-}S\}$

$D \in \{I', II', III', IV', IV\text{-}S'\}$

Also $P[X_t=i]$ and $P[X_t=i']$ were developed as functions of probability elements in the last section.

In order to make the O.F. an explicit function of Z_t, the probability elements a_{ij}, b_i and U_i's are substituted by their equivalent values, i.e.,

$$a_{ijt} = a'_{ijt}[1-Z_t \cdot f_i(t)] \qquad i \neq 0$$

$$U_{it} = U'_{it}[1-Z_t \cdot f_i(t)] \qquad i \neq 0$$

$$b_{it} = b'_{it}[1-Z_t \cdot f_i(t)] + Z_t \cdot f_i(t)$$

This reduces the objective function to only a function of Z_t, $1 \leq t \leq t_{max}$. If the true positive rate of the test is taken to be a constant independent of age and stage, then the analysis of Section 3.5 could be employed to write the objective function in a more compact form as,

$$O.F. = \sum_{T_j \in T} \left(e^{-\lambda_o T_{j-1}} - e^{-\lambda_o T_j}\right) \left\{ \sum_{t=T_{j-1}+1}^{t_{max}} (1-F)^{\sum_{k=T_{j-1}+1}^{t} Z_k} \cdot \right.$$

$$\cdot \left[\sum_{i'} Z_t \cdot C_s(t) \cdot \frac{P'[X_{t+1}=i']}{b'_{i,t+1}} \right] + (1-F)^{\sum_{k=T_{j-1}+1}^{t-1} Z_k}$$

$$\left. \left(1-F+\frac{F}{b'_{i,t}}\right)^{Z_t} \cdot C_{t,i'} \cdot P'[X_t=i'] \right\} + \sum_{t=1}^{t_{max}} Z_t \cdot C_s(t) \cdot$$

$$P[X_t=0]$$

where $C_{t,i'} = C_T(t,i') + C_d(t) \cdot d'_i(t)$. The optimization problem can be solved by enumerating all possible cases - by evaluating the value of

O.F. using the policy under consideration and then selecting the policy which gives minimum expected total cost of screening.

4.4 Parameter Estimation and Determination of Optimal Policies

The cost-effectiveness measure is not always the best measure to use, but it gives an idea of how much cost will be saved by optimizing the number of screens. A major controversy over the use of the cost-effectiveness measure results from associating a fixed cost to the death of an individual. Determination of this cost is very difficult and depending on the analyst's point of view the value varies drastically. The U. S. Department of Health, Education and Welfare has published a book [80] on the estimation of the cost of illness and has developed an estimate of the present worth of the lifetime earnings for males and females using 4% and 6% rates on the data for year 1963. Table 4 shows the present worth of the future earnings. These data were used as the cost of death at any age t.

The range of ages of interest for Neuroblastoma is between 0 and 10 years and, as shown in Figure 10, in this range the cost of death $C_d(t)$ behaves linearly. Therefore, a least square model is fitted through the first seventeen years.

$$Y^T = [59063 \quad 64789 \quad 79333 \quad 96736 \quad 114613]$$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 25 & 2.5 & 7 & 12 & 17 \end{bmatrix}$$

where Y^T is the cost vector and X^T is the age matrix.

Table 4: Estimation of the present worth of the lifetime earning (extracted from reference 80).

	Male		Female	
AGE	4%	6%	4%	6%
Under age 1	59063	30622	34622	18189
Age 1 - 4	64989	34998	37938	20708
5 - 9	79333	46986	46289	27791
10 - 14	96736	62994	56422	37240
15 - 19	114613	80953	64936	45990
20 - 24	126688	94811	67960	49987
25 - 29	128698	100011	66828	50205
30 - 34	122904	98298	64389	49367
35 - 39	111956	91864	60998	47744
40 - 44	97301	81766	56603	45286
45 - 49	80325	69025	50896	41568
50 - 54	63027	55410	44371	36961
55 - 59	45948	41334	37467	31845
60 - 64	28387	25970	30164	26126
65 - 69	15043	13737	23579	20823
70 - 74	9264	8573	18118	16406
75 - 79	5344	5014	12888	12024
80 - 84	2935	2791	6916	6654
85 Over	210	205	1123	1096

NOTE: It is possible to find a common cost for undetermined sex by mixing the two respective costs by the male/female ratio.

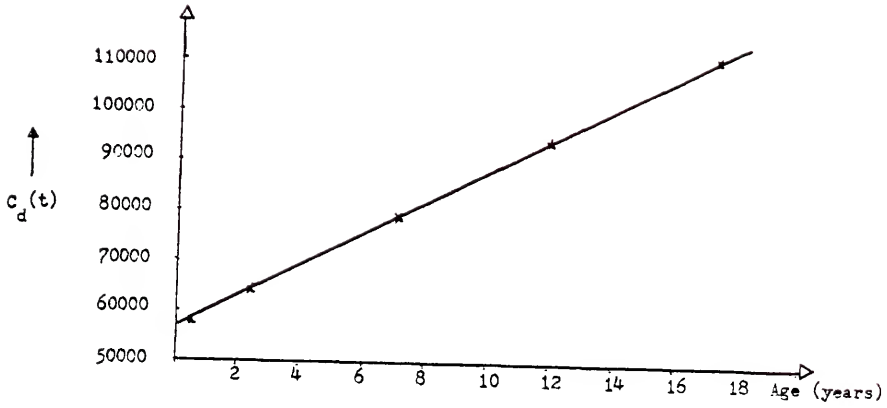


Figure 10: Least square linear model fitted through data on $C_d(t)$.

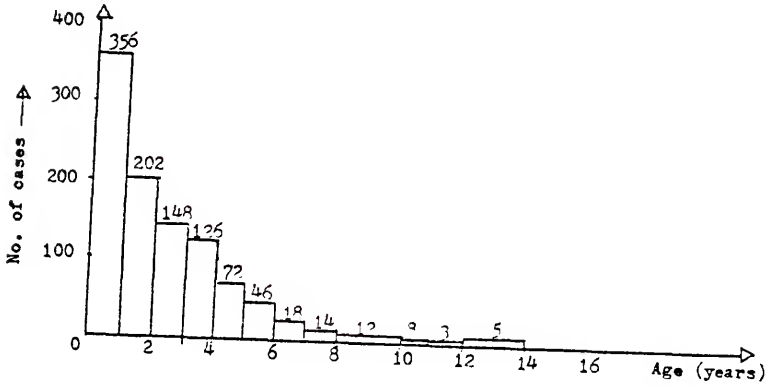


Figure 11: Age distribution of 1027 children with Neuroblastoma.

The slope of the line is,

$$\beta = (X^T X)^{-1} (X^T Y) = (59{,}558.97 \quad 3{,}376.19)^T$$

and the relationship as a function of time is,

$$C_d(t) = 56{,}558.97 + 3{,}376.19t$$

where t is in years.

In order to get a form for $P[X_t=i']$, the data gathered by Voute et al. [105] are used to provide the age distribution at diagnosis for a total of 1027 cases. Figure 11 shows the age distribution of 1027 children with Neuroblastoma.

The data seem to be generated by a negative exponential distribution. Therefore, in order to estimate the expected age at diagnosis the following analysis is used,

$$E[\text{Age of diagnosis}] = \sum_{t=1}^{14} N_t \cdot t / \sum_{t=1}^{14} N_t = 2.9883 \text{ years}$$

$$\lambda' = \text{Rate of detection} = \frac{1}{E[\text{Age at diagnosis}]} = 0.3346 \text{ years}$$

To see the validity of the assumption of a negative exponential distribution for the age at detection, the empirical distribution is compared with the theoretical one in Table 5.

The data presented do not specifically determine the probability of being in any state i'. Instead, they indicate the probability of being in the detected state at time t. In order to find the probability of being in any specific detected state i' it is necessary to have additional information on the way this quantity is distributed between different

Table 5: Comparison of empirical and theoretical distribution.

t	1	2	3	4	5	6	7	8	9	10
F(t)	.284	.487	.633	.737	.812	.865	.903	.931	.95	.946
$\sum_{i=1}^{t} n_i$	292	501	650	757	834	889	928	956	976	990
n_t	292	209	149	107	77	55	39	28	20	14
data	356	202	148	126	72	46	18	14	12	12

stages at any age of interest. This information is not presently available and therefore in this analysis it was taken to be equiprobable in any of the five stages I, II, III, IV and IV-S (irrespective of the age of the individual). Later, different distributions are used to check the sensitivity of the results using different distributions.

To compute the probability of death as a function of age and stage $d_i(t)$, model D developed by Breslow and McCann [13] is used. This model has been referred to by many other investigators. It says,

$$\ell n \frac{P_t}{1 - P_t} = a_i + d \cdot \ell n(t+6)$$

where P_t is the probability of survival at age t; t is the age at diagnosis in months; a_i, i=I,II,III,IV,IV-S, and d are parameters of the model whose values are estimated to be,

$$a_I = 6.08$$

$$a_{II} = 4.77$$

$$a_{III} = 4.07$$

$$a_{IV} = 1.73$$

$$a_{IV\text{-}S} = 5.15$$

$$d = -1.36$$

This reveals that,

$$d_i(t) = 1 - [1+\exp(-(a_i+d\cdot \ell n(t+6)))]^{-1}$$

The incidence rate of Neuroblastoma is not known accurately but the range of its variation is between (1/30,000) to (1/100,000). In the analysis the incidence rate was taken to be (1/30,000) children, which means one out of every 30,000 children will be involved with this disease.

The program used in analysis performs the following tasks:

1) It generates different policies with the following constraint on the number of tests done and the interval between them:

a) Total number of tests done for an individual is less than or equal to three.

b) Minimum interval between two screening tests is,

1 month for $0 \leq t \leq 24$ months

6 months for $24 < t < 48$ months

12 months for $t > 48$ months.

2) It computes and prints different elements of cost for each policy.

3) It finds the best policy.

4.5 Sensitivity Analysis

Due to the fact that Neuroblastoma is not a well-known disease-- its structure, progression, onset rate, cure rate and regression are not completely known to date-- it is impossible to use a single number to

represent any of the above mentioned quantities and find the best policy for screening the disease. Therefore, it remains to vary the unknown parameters over their entire range and find the optimal policy under any of those situations. Then it might be possible to find certain patterns and/or to show that policy selection is not sensitive to a particular parameter.

In the program the following values are used as the best estimate of their corresponding quantities,

1) Treatment cost for an individual who is diagnosed in state I = \$7,000

Treatment cost for an individual who is diagnosed in state II = \$8,000

Treatment cost for an individual who is diagnosed in state III = \$9,000

Treatment cost for an individual who is diagnosed in state IV = \$10,000

Treatment cost for an individual who is diagnosed in state IV-S = \$7,500.

2) The screening cost per individual is the cost of one VMA strip test which according to Leonard et al. [64] is 15¢.

3) The true positive rate of VMA test is not known accurately, but its range of variation is believed to be $0.70 \leq F \leq 0.9$ [44,46,64,108]. Therefore, in the computation, different runs are made by varying F in steps of 0.10 over its entire range.

4) It was mentioned before that to give an estimate of a_{ij}, U_i and b_i's, the system of 11 equations and 16 unknowns should be solved for each time interval. It is clear that 5 of the unknowns must be estimated

using some other methods. Temporarily, it was decided not to use the procedure of estimating a_{ij}, b_i and U_i's, but to give a value to b and find the optimal policy by varying b over a reasonable range. By definition,

$$b_{i,t+1} = \frac{P[X_{t+1} = i']}{P[X_t = i]}$$

In the case of Neuroblastoma there is no idea what the duration of stay in the occult stages is; therefore, in the program a range of b' = 0.04 to b' = 0.07 was used, which means duration of stay in the occult stages varies from 15 months to 25 months (these are values of the detection rate under the natural progression of the disease).

5) There is no report on the distribution of onset of the Neuroblastoma ; there are some reports presenting data and claiming that the data is on the age at onset but in fact it is the age at detection. Therefore it was decided to assume the onset of the disease is negative exponential distributed with rate λ_0. Then by varying λ_0 from 0.07 to 0.15 in steps of 0.02 the sensitivity of the optimal solution versus λ_0 was investigated. To see the effect of varying λ_0 on the progression of the disease, the probability of being free of disease at age t is computed for two different cases - only those individuals are considered that eventually will get the disease,

For $\lambda_0 = 0.03$ $P[X_{12}=0] = 0.697$, $P[X_{24}=0] = 0.486$, $P[X_{36}=0] = 0.339$

For $\lambda_0 = 0.06$ $P[X_{12}=0] = 0.486$, $P[X_{24}=0] = 0.236$, $P[X_{36}=0] = 0.115$

Table 6 shows the output of computer runs including different costs of the optimal policy for different values of parameters.

4.6 Results and Conclusions

The following primary conclusions can be obtained by looking at the graphs:

1) Table 6 shows that assuming that the cost of one test per individual is fifteen cents, it is always optimal to have the child tested at least once during his life. The time of screening varies with parameters but it is mostly during the second year of life.

2) The optimal solution is not very sensitive to F, the true-positive rate of test. In fact, increasing F decreases the cost of screening but mostly does not change the time of screening. Figure 12 shows the effect of F on the cost of screening for different values of onset rate and conditional probability of detection. Insensitivity of the optimal policy to F shows that in the case of Neuroblastoma, the assumption of constant F-- irrespective of age and stage-- is a relatively valid assumption.

3) Increasing λ_0 (onset rate) decreases the expected total cost of screening and shifts the optimal screening time toward date of birth. This is because higher values of λ_0 correspond to earlier involvement with the disease and a higher probability of survival if detection occurs. Figure 13 shows the effect of onset rate on total cost of screening for different values of F and b.

4) Increasing b (the conditional probability of detection) decreases the total cost of screening and the optimal policy tries to postpone the time of screening until the child gets older. This is because of the fact that higher values of b correspond to better natural detection of the disease which reduces the benefits from screening. Figure 14 shows the effect of conditional probability of detection on total cost for different values of F and λ_0.

Table 6: The cost of the optimal policy for different values of the model parameters.

λ	b	total cost F=0.7	Policy No.	total cost F=0.8	Policy No.	total cost F=0.9	Policy No.
0.07	0.04	30594	5	29951	5	29308	5
	0.05	29143	6	28304	6	27465	6
	0.06	27951	35	26986	35	26131	35
	0.07	26959	35	25907	35	24967	36
0.09	0.04	30607	5	29860	5	29113	5
	0.05	29044	5	28074	5	27104	5
	0.06	27766	34	26724	35	25764	5
	0.07	26683	35	25521	35	24508	35
0.11	0.04	30552	4	29721	4	28889	4
	0.05	28883	5	27806	5	26729	5
	0.06	27514	34	26434	34	25243	5
	0.07	26346	34	25177	34	24146	26
0.13	0.04	30354	4	29442	4	28529	4
	0.05	28711	4	27563	4	26416	4
	0.06	27216	25	26009	25	24833	5
	0.07	26077	25	24775	25	23654	25
0.15	0.04	30189	4	29208	4	28228	4
	0.05	28424	4	27191	4	25958	4
	0.06	26877	25	25623	25	24445	4
	0.07	25679	25	24328	25	23176	25

Notation : Policy no. 4 = One screen at age 9 months.
" " 5 = " " " " 12 "
" " 6 = " " " " 15 "
Policy no. 25=Two screens at ages 6 and 15 months.
" " 26= " " " " 6 " 18 "
" " 34= " " " " 9 " 18 "
" " 35= " " " " 9 " 21 "
" " 36= " " " " 9 " 24 "

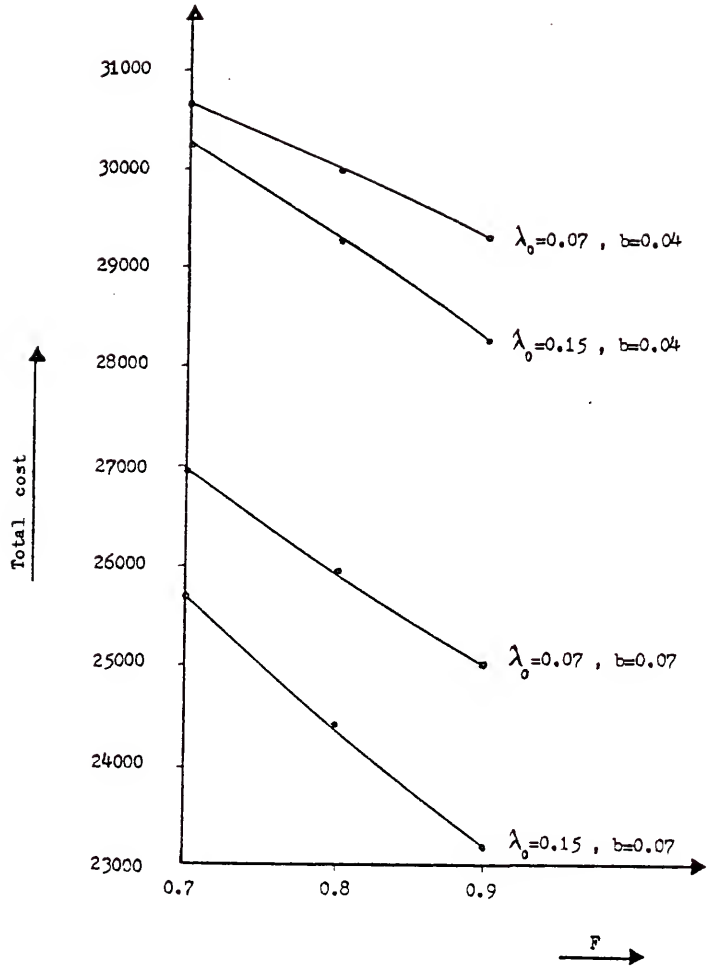


Figure 12: The effect of true positive rate of test on total cost.

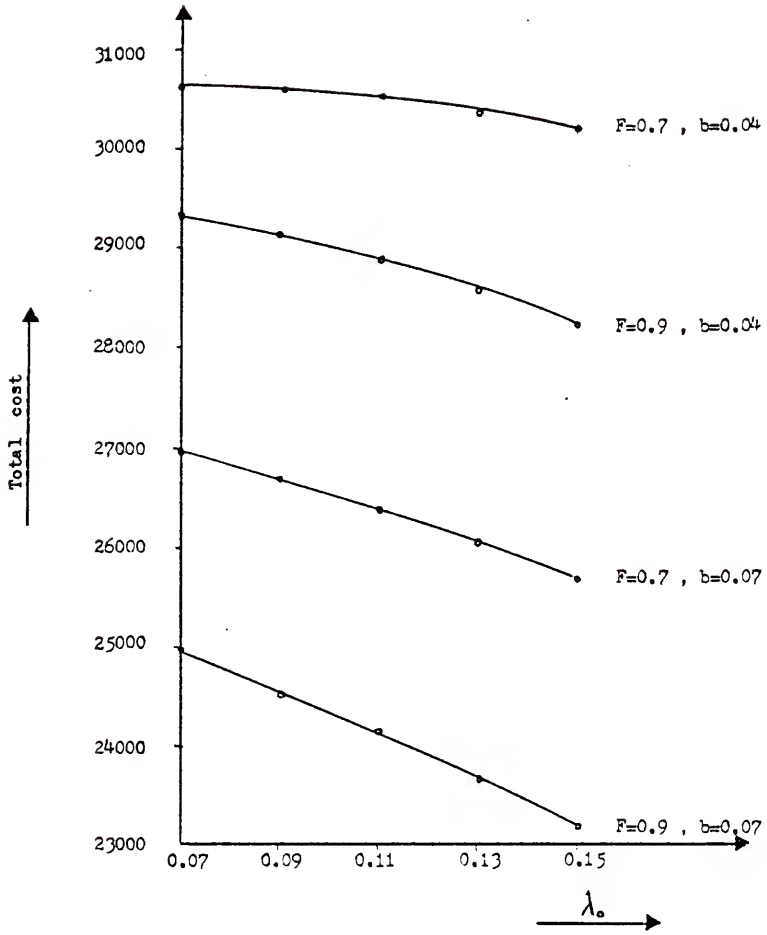


Figure 13: The effect of the onset rate on the total cost of screening.

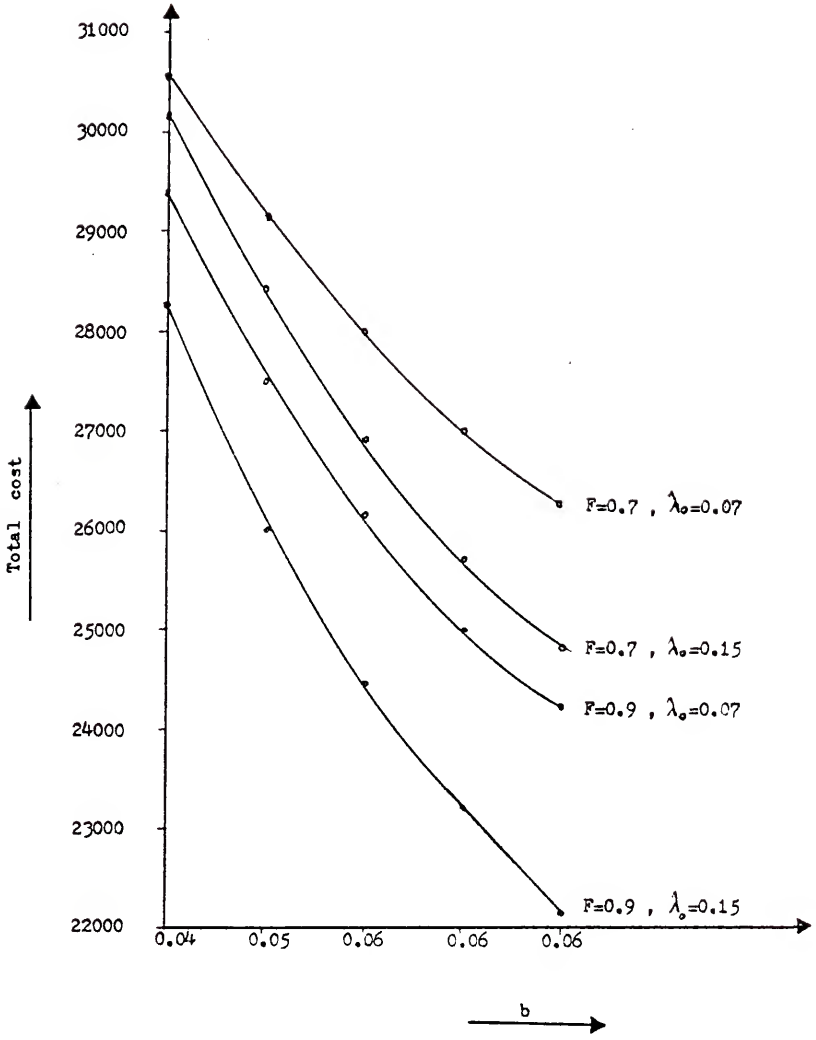


Figure 14: The effect of the conditional probability of detection on the total cost of screening.

5) The general format of the different elements of screening costs as a function of time of the test is shown in Figure 15. It is seen that due to the chosen estimates of costs of treatment and death, the only major element is cost of death at age t which governs the shape of the curve. In general, expected total cost of screening decreases up to age t and increases afterwards. The curve is for a single test and could be modified for the case of more than one test per lifetime.

6) Figure 16 shows the sensitivity of optimal cost to variation of different parameters. It is seen that,

a) Increasing b is generally associated with lower estimate of total cost.

b) Increasing λ_0 decreases the cost of screening up to age t and increases it afterwards. This is because if everybody gets the disease early in life, it is not optimal to screen them late in life when the disease has progressed to its final stages.

7) Table 7 shows the optimal screening age and associated cost for a fixed number of screens. It is used to decide on the future screening of a child who has already been screened once. It gives the best choice of future time of screening as a function of number of desired screenings.

It is seen that in this particular case it is optimal to screen the child twice, at ages 9 and 21 months.

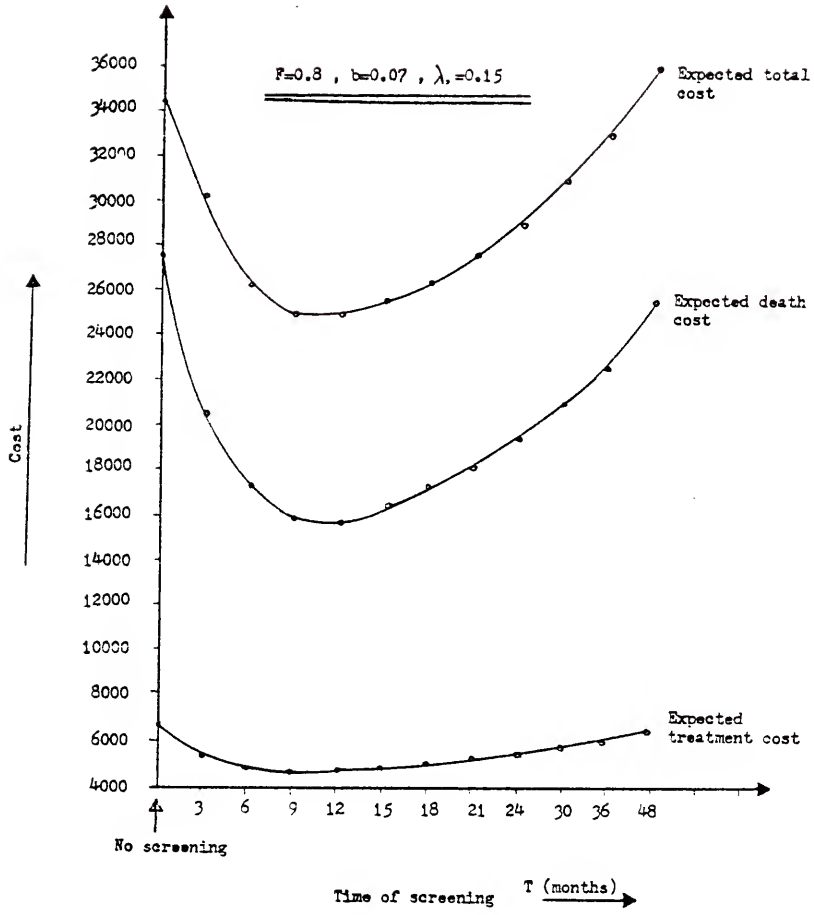


Figure 15: The elements of screening cost for a single screening at month T.

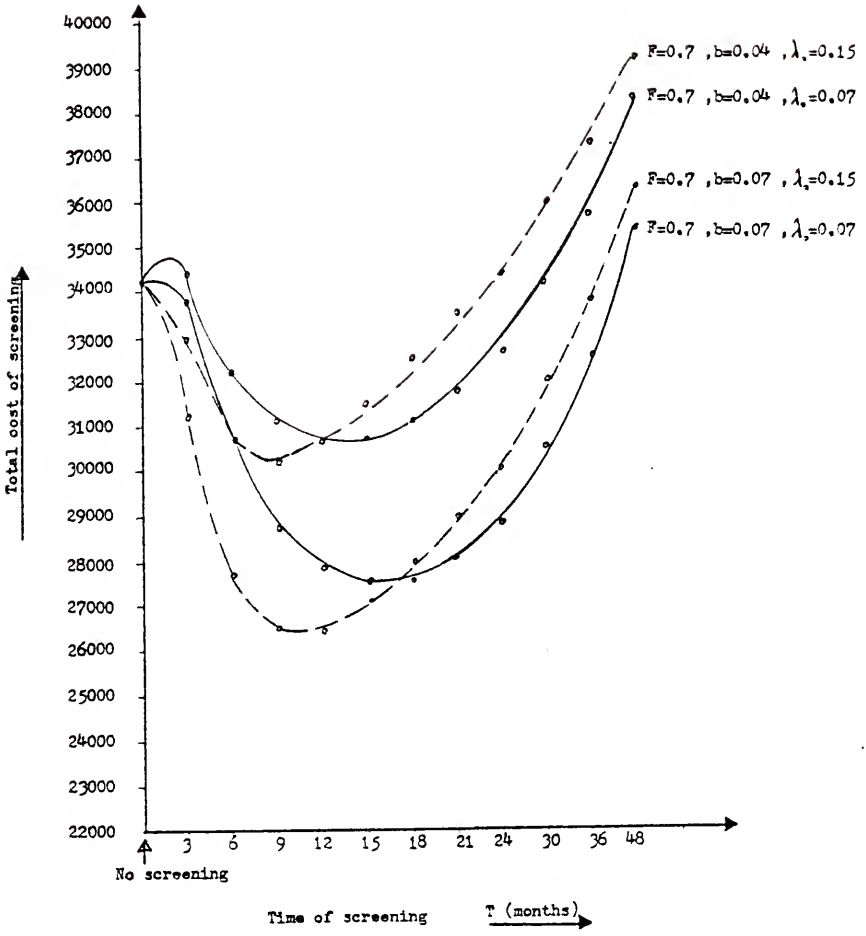


Figure 16: Comparison of the total cost of a single screening at different ages.

Table 7: Cost of optimal screening for a fixed number of screens.

For λ_0=0.07 , b =0.06 , and F =0.8 the optimal cost of screening under following conditions is :

A - Having no screen : 34247 $

B - Having one screen :27152 $ (at age 15 months)

C - Having two screens:26986 $ (at ages 9 and 21 months)

D - Having three screens : 29009 $ (at ages 6 , 15 and 30 months)

For two screens ;

If the first one is done at :	The second one better be at :	The associate cost is :
3	18	28913
6	18	27500
9	21	26986 *****
12	24	27162
15	24	27837
18	30	28749
21	30	29818
24	30	31033
30	36	33430

For three screens :

If the first one is done at :	The second and third one better be at :	The associate cost is :
3	12 , 24	29642
6	15 , 30	29009 *****
9	18 , 30	29016
12	18 , 30	29613
15	21 , 30	30590
18	24 , 36	31709
21	24 , 36	33027

CHAPTER FIVE

SPECIAL CASES

In this chapter a variety of subjects related to reformulation of the problem under different conditions is considered. In fact, the main assumptions of the model are altered and a sensitivity analysis is employed to determine the robustness of the model relative to changes in some of the parameters. In the first three sections the assumption of constant true positive rate of screening, used in arguments up to now, is relaxed and the general expression for expected total cost of screening is derived for each of the following cases:

1) The true positive rate of screening on two successive examinations are dependent.

2) The true positive rate of screening is a function of preclinical time which is time from the onset of the disease.

3) Each screening examination consists of a sequence of tests and there is a transition from false positive groups to true positive groups through time.

Up to this point, all derivations have been based on the assumption of steady state condition. Section 5.4 investigates the situation in which a new screening program has just started (as a function of time from establishment of the program), and the expression for expected total cost of screening is derived.

In Section 5.5 the model is investigated for the case of breast cancer. To do so, data from the control group of the HIP study on breast

cancer is used as input and the model's output is compared with that of the study group.

5.1 Dependency of True Positive Rate of Two Successive Examinations

In the case of some diseases, it is possible that if a diseased individual is screened for the first time and gives a false negative response to the examination, he has a higher chance of responding negatively from then on. Therefore a more difficult question, perhaps, is that of the relationship between the false negative (or equivalently true positive) probability on successive examinations of the same individual. The case of complete independence was derived by letting F be a constant independent of how many times the individual has falsely responded negative. Under that assumption, if the probability that the disease is missed on one examination is (1-F), then the probability that it is missed on M screens would be $(1-F)^M$. In this section there will be a correlation between the probability of true positive in two successive screening examinations.

Assume that true positive incidence rate of the (i+k)-th screen, given that the individual, who got the disease at or before T_i, was not detected at T_i, $T_{i+1}, \ldots T_{i+k-1}$ be F_{k+1}.

This means a diseased individual will respond to the sequence of tests with probabilities F_1, F_2, $F_3, \ldots$. To develop a cost-effectiveness objective function, consider the first group--those who have had the disease by the time of the first screening. For this group, at any time t in the interval:

$$0 \leq t < T_1 : \qquad P[X_t=i] = P'[X_t=i]$$
$$P[X_t=i'] = P'[X_t=i']$$

At $t=T_1$:

$$P[X_{T_1}=i] = (1-F_1)\cdot P'[X_{T_1}=i]$$

$$P[X_{T_1}=i'] = (1-F_1+F_1/b_{T_1})\cdot P'[X_{T_1}=i']$$

This is because at $t=T_1$, the individual has participated in the first examination. For t in the interval $T_1 < t < T_2$:

$$P[X_t=i] = (1-F_1)\cdot P'[X_t=i]$$

$$P[X_t=i'] = (1-F_1)\cdot P'[X_t=i']$$

At time T_2 :

$$P[X_{T_2}=i] = (1-F_2)(1-F_1)\ P'\ [X_{T_2}=i]$$

$$P[X_{T_2}=i'] = (1-F_1)(1-F_2+F_2/b_{T_2})\ P'[X_{T_2}=i']$$

This is because the remaining individuals $(1-F_1)\cdot P'[X_t=i]$ will be detected with rate F_2 and $(1-F_2)(1-F_1)P'[X_{T_2}=i]$ will remain in the occult state.

The same type of analysis holds through for $T_2 < t < T_3$ and can be generalized to establish that for the first group the contribution to the cost is

$$\sum_{t=1}^{t_{max}} \sum_{i} \left\{ \left[\prod_{k=1}^{\sum_{\ell=1}^{t} Z_\ell} (1-F_k) \right] \cdot Z_t \cdot C_s(t) \cdot \frac{P'[X_{t+1}=i']}{b'_{i,t+1}} + \right.$$

$$\left. \left[\prod_{k=1}^{\sum_{\ell=1}^{t-1} Z_\ell} (1-F_k)\ (1-F_k + \frac{F_k}{b'_{i,t}})^{Z_t} \right] \cdot C_{ti'} \cdot P'[X_t=i'] \right\}$$

For the second group--those who get the disease at sometime between T_1 and T_2--the analysis is

For $T_1 < t < T_2$:

$$P[X_t=i] = P'[X_t=i]$$

$$P[X_t=i'] = P'[X_t=i']$$

at $t=T_2$:

$$P[X_{T_2}=i] = (1-F_1)P'[X_{T_2}=i]$$

$$P[X_{T_2}=i'] = (1-F_1+F_1/b_{i,T_2})P'[X_{T_2}=i']$$

This is because at this time they undergo their first screening examination.

Considering the contribution of all the groups, the general expression for the expected total cost of screening is

$$\text{O.F.} = \sum_{j=1}^{M+1} \left\{ \sum_{t=T_{j-1}+1}^{t_{max}} W(j,t) \left\{ \sum_{i'} \left(\left[\prod_{u=1}^{\sum_{\ell=T_{j-1}+1}^{t} Z_\ell} (1-F_u) \right] \cdot Z_t \cdot C_s(t) \cdot \frac{P'[X_{t+1}=i']}{b'_{i',t+1}} \right) + \right.\right.$$

$$\left.\left. \left(\left[\prod_{u=1}^{\sum_{\ell=T_{j-1}+1}^{t-1} Z_\ell} (1-F_u)(1-F_u+\frac{F_u}{b'_{it}})^{Z_t} \right] \cdot C_{ti'} \cdot P'[X_t=i'] \right) \right\} \right\}$$

$$+ \sum_{t=1}^{t_{max}} Z_t \cdot C_s(t) \cdot \{P[X_t=0]+I\}$$

where I = number of population per one patient.

5.2 True Positive rate as a Function of Time from Onset of the Disease

In the case of some progressive diseases such as breast cancer, the true positive rate of screening test could be a function of the tumor size, which is a function of preclinical time. In such a case, it is reasonable to assume true positive rate of a test is a function of the time that individual has had the disease.

In order to bring this generalization into account, the model should be modified, because of the particular structure of the model (there are M+1 groups, the i-th group of which has an onset time sometime between (i-1)th and i-th examination) there is no problem in generalization of the expression. To see this for the first group, the true positive rate would be f(t) where t is age, whereas for the second group, it would be $f(t-T_1)$ and for the third group $f(t-T_2)$ and so forth. The general expression would simply be

$$\sum_{j=1}^{M+1} \left\{ \sum_{t=T_{j-1}+1}^{t_\infty} W(j,t) \left(1-f(t-T_{j-1})\right) \sum_{k=T_{j-1}+1}^{t} Z_k \left[\sum_i Z_t \cdot C_s(t) \cdot P'[X_t=i] \right] + \right.$$

$$\sum_{t=T_{j-1}+1}^{t_\infty} W(j,t)\,[1-f(t-T_{j-1})] \sum_{k=T_{j-1}+1}^{t-1} Z_k \left[\sum_{i'} \left(1-f(t-T_{j-1}) + \frac{f(t-T_{j-1})}{b'_{it}}\right) Z_t \right.$$

$$\left. \left. \cdot\, C_{ti'} \cdot P'[X_t=i'] \right] \right\} + \sum_{t=1}^{t_{max}} Z_t \cdot C_s(t) \cdot \{P[X_t=0]+I\}$$

In order to see the effect of variation of F with time from onset, a computer program was run for the following functional forms of F:

Case 1: $f(t) = F_0 + a.t.$ for $F_0 = 0,.2,.4,.6,.8$

Case 2: $f(t) = F_0 + a.t.^2$ for $F_0 = 0,.4$

Case 3: $f(t) = 1-\exp(-\lambda t)$ for $\lambda = 1/12$

The comparison between different cases can be seen from Table 8 in which the optimal screening policy along with the associated expected total cost is presented.

Table 8: Comparison between different cases in optimal policy and associated cost.

f(t)	0.7	t/72	.2+t/40	.4+t/120	.6+t/180	.8+t/240	$t^2/5184$	$.4+t^2/8640$	$1-e^{-\frac{t}{12}}$
time of the optimal test (in months)	12	12	12	12	12	12	12	12	9
optimal cost (in dollars)	30425	30969	30474	29980	29485	28991	32445	30865	28707

Figures 17 and 18 show the expected total cost as a function of time of screening for the case of one screening per life time for cases 1, 2 and 3. It is seen that there is no significant change in the optimal policy but the expected total cost decreases as F_0 increases.

It can be concluded that in the case of Neuroblastoma, as far as determination of the optimal policy is concerned, the assumption of a constant true positive rate of screening is a reasonable one.

5.3 Screening Examination Consists of a Sequence of Tests

In a mass screening program the primary test is not usually the most sophisticated test used, due to time limitations and cost considerations. Therefore an individual who shows a positive response is only a potential carrier of the disease and a more sophisticated series of tests is required before the individual begins treatment. In most cases a person who responds positively to the primary examination, but negatively to the more specific sequences of tests, is considered a high risk individual and might be urged to be screened more often.

Under these circumstances it is interesting to see how the elements of cost differ from the simpler main model. The model should be related to the dynamics of the disease in the individual and should be modeled as a function of time. Any reasonable assumption on the dynamic behavior of the disease in an individual could be employed to build the model. In what follows, one possible set of assumptions is mentioned. Based on these assumptions, a model is developed which considers an individual who starts the disease with certain probability of being detected by a scheduled examination. As time passes his chances of detection increase with certain rates until he is clinically surfaced.

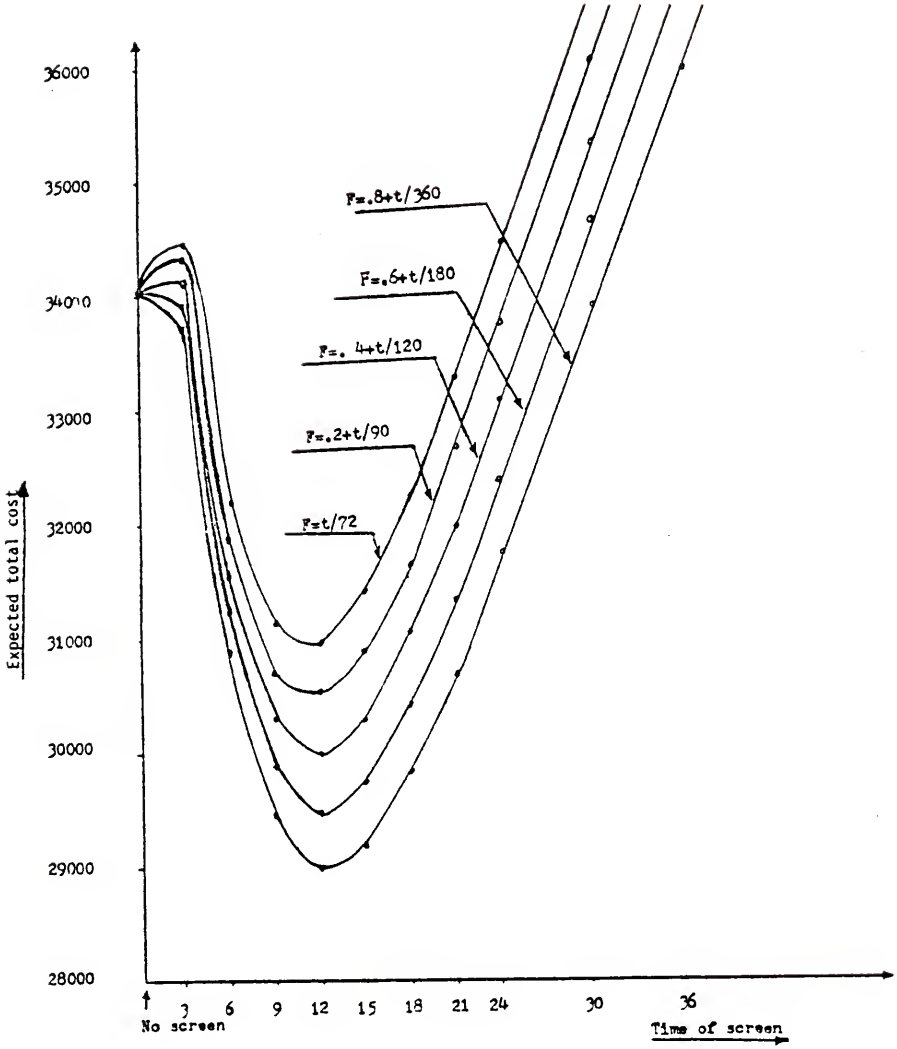


Figure 17: Comparison for the case of one screen per lifetime.

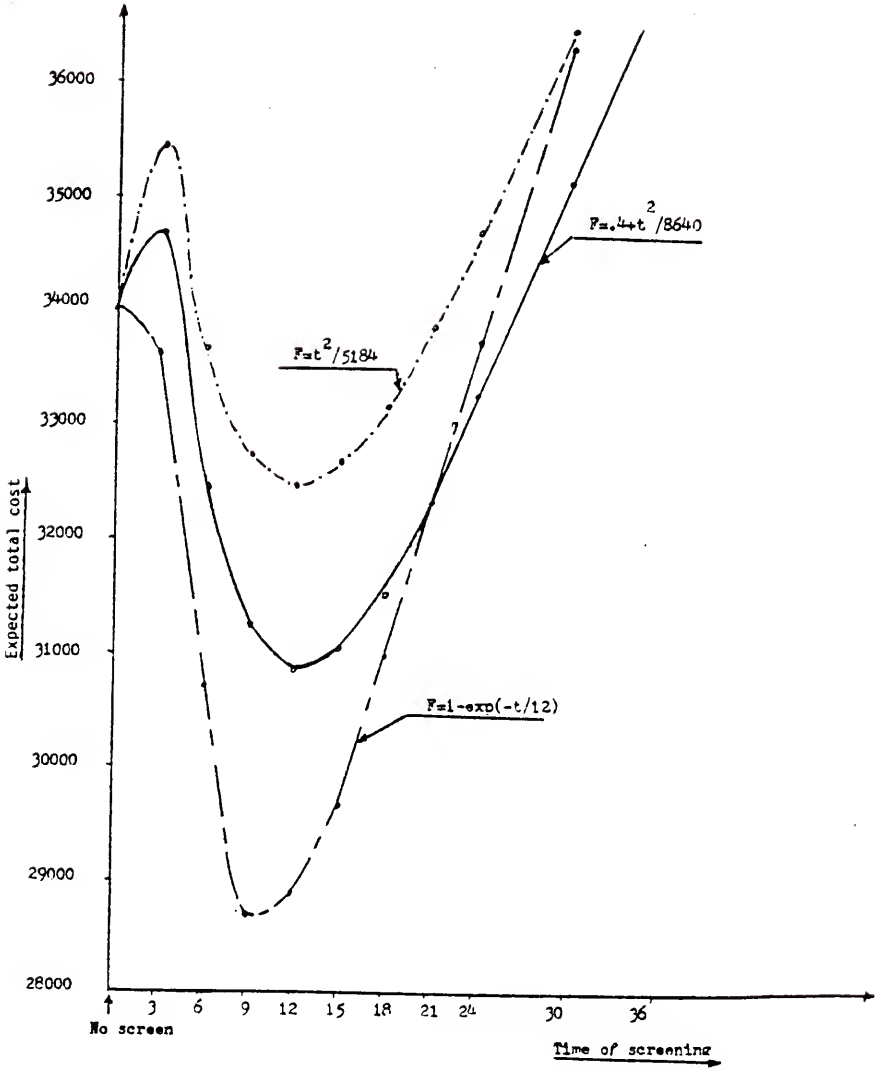


Figure 18: Comparison for the case of one screen per life time.

In order to analyze the model the following notation is employed:

a) An individual who will never get the disease is denoted by $(-)_d$.

b) An individual who will eventually get the disease is denoted by $(+-)_d$.

c) An individual who has the disease is denoted by $(++)_d$.

d) An individual who responds negatively to all examinations is denoted by $(-)_s$.

e) An individual who responds positively but reveals no evidence of the disease in subsequent tests is denoted by $(+-)_s$.

f) An individual who responds positively to all the tests is denoted by $(++)_s$.

To clarify the notation, Figure 19 is presented. At any particular time t an undetected--probably healthy--individual has to be in a particular block with respect to the nature of the disease [$(-)_d$, $(+-)_d$, $(++)_d$] and the result of the screen [$(-)_s$, $(+-)_s$, $(++)_s$]. For instance an individual in block C carries the disease at time t but can't be detected through screening. The values in each block represent the ratio of the population in each category at any particular time t.

The assumptions are as follows

1) A $(-)_d$ individual will always give the same response to a test. I.e., if he responds $(-)_s$, he will respond $(-)_s$ to all screening examinations of the same type in his lifetime.

2) For a $(-)_d$ individual the probability of being in any of the $(-)_s$, $(+-)_s$ and $(++)_s$ categories is $1-(F_1+F_2)$; F_1; and F_2 respectively, where F_1 and F_2 are constants.

		The nature of the disease		
		$(-)_d$	$(+-)_d$	$(++)_d$
Result of a screen	$(-)_s$	A $I \cdot [1-(F_1+F_2)]$	B $P_0[1-(F_3+F_4)]$	C $P_t[1-(F_5+F_6)]$
	$(+-)_s$	D $I \cdot F_1$	E $P_0 \cdot F_3$	F $P_t \cdot F_5$
	$(++)_s$	G $I \cdot F_2$	H $P_0 \cdot F_4$	K $P_t \cdot F_6$
Total		Population incidence	$P_0=P[X_t=0]$	$P_t=P[X_t=d]$

Figure 19: Categorization of population into 9 groups--A through K--according to the nature of the disease and result of the screen.

3) A $(+-)_d$ individual will always give the same response to the test.

4) For a $(+-)_d$ individual let the probability of responding $(-)_s$, $(+-)_s$, and or $(++)_s$ be $1-(F_3+F_4)$; F_3; and F_4 respectively, where F_3 and F_4 are constants.

5) A $(++)_d$ individual starts the disease in one of the $(-)_s$; $(+-)_s$; and $(++)_s$ categories. Corresponding probability is $1-(F_5+F_6)$; F_5; and or F_6 respectively. Moreover

5.1) A $(-)_s$ individual may switch to $(+-)_s$, (or $(++)_s$) category with probability $\lambda_1 t$, ($\lambda_2 t$), where t is time from onset. Therefore this individual has the following probabilities of being in $(-)_s$, $(+-)_s$, or $(++)_s$ categories:

$$1-(F_5+F_6)-(\lambda_1+\lambda_2)\cdot t \; ; \; \lambda_1 t; \; \lambda_2 t$$

5.2) A $(+-)_s$ individual may switch to $(++)_s$ category with probability $\lambda_3 t$, where t is time from onset. Therefore this individual has the following probabilities of being in $(+-)_s$ and $(++)_s$ categories;

$$F_5 + \lambda_1 t - \lambda_3 t \; ; \; \lambda_3 t$$

Putting all the assumptions together, the following observations are made:

1) The probability of a response being $(-)_s$ as a function of time is

$$R_1 = 1-(F_5+F_6) - (\lambda_1+\lambda_2)\cdot t$$

$$= 1-(F_5+F_6) - \underbrace{\lambda_1 t\left[1- \frac{\lambda_3 t}{F_5+\lambda_1 t}\right]}_{\text{this goes to } (+-)_s} - \underbrace{\left[\lambda_2 t + \lambda_1 t \left(\frac{\lambda_3 t}{F_5+\lambda_1 t}\right)\right]}_{\text{this goes to } (++)_s}$$

ii) The probability of a response being $(+-)_s$ as a function of time is

$$R_2 = F_5 + \lambda_1 t - \lambda_3 t$$

$$= F_5 + \underbrace{\lambda_1 t \left[1 - \frac{\lambda_3 t}{F_5+\lambda_1 t}\right]}_{\text{this comes from } (-)} - \underbrace{\lambda_3 t \left[1 - \frac{\lambda_1 t}{F_5+\lambda_1 t}\right]}_{\text{this goes to } (++)}$$

(iii) The probability of a response being $(++)_s$ as a function of time is

$$R_3 = F_6 + (\lambda_2+\lambda_3)\, t$$

$$= F_6 + \underbrace{\left[\lambda_2 t + \lambda_1 t \left(\frac{\lambda_3 t}{F_5+\lambda_1 t}\right)\right]}_{\text{this comes from } (-)_s} + \underbrace{\lambda_3 t \left[1 - \frac{\lambda_1 t}{F_5+\lambda_1 t}\right]}_{\text{this comes from } (+-)_s}$$

Note that $R_1+R_2+R_3=1$, which means an individual has to be in one of the three mutually exclusive categories.

To find an expression for the cost-effectiveness objective function, the following analysis is developed. For the first group, who have the disease by the time of the first screen, for

$$0 < t < T_1 \; : \; P[X_t=i'] = (R_1+R_2+R_3)\, P'[X_t=i'] = P'[X_t=i']$$

$$\text{at } t=T_1 \; : \; P[X_{T_1}=i'] = \left[(R_1+R_2) + \frac{R_3}{b_{iT_1}}\right] \cdot P'[X_{T_1}=i']$$

$$= \left\{[1-F_6-(\lambda_2+\lambda_3)T_1] + \frac{F_6+(\lambda_2+\lambda_3)T_1}{b_{iT_1}}\right\} \cdot P'[X_{T_1}=i']$$

$$= \left\{1-[F_6+(\lambda_2+\lambda_3)T_1]\left(1 - \frac{1}{b_{iT_1}}\right)\right\} P'[X_{T_1}=i']$$

Of course at time T_1--just after the test is done--the $(++)_s$ category

becomes empty but it gradually builds up by those switching from $(-)_s$ and $(+-)_s$ categories. At any time t this pool contains $S_1=(\lambda_2+\lambda_3)(t-T_1)$ individuals in $(++)_s$. Therefore for $T_1 < t < T_2$

$$P[X_t=i'] = (R_1+R_2+S_1)\ P'[X_t=i']$$

$$= \left\{[1-F_6-(\lambda_2+\lambda_3)t] + (\lambda_2+\lambda_3)(t-T_1)\right\} P'[X_t=i']$$

$$= \left\{1-[F_6+(\lambda_2+\lambda_3)T_1]\right\} \cdot P'[X_t=i']$$

At $t=T_2$

$$P[X_{T_2}=i'] = (R_1+R_2+S_1/b_{iT_2}) \cdot P'[X_{T_2}=i']$$

$$= \left\{[1-F_6-(\lambda_2+\lambda_3)T_2] + (\lambda_2+\lambda_3)(T_2-T_1)/b_{iT_2}\right\} \cdot P'[X_{T_2}=i']$$

$$= [1-F_6-(\lambda_2+\lambda_3)(T_2-(T_2-T_1)/b_{iT_2})] \cdot P'[X_{T_2}=i']$$

Once more at time T_2--just after the test is done--the $(++)_s$ category becomes empty, but it builds up by those switching from $(-)_s$ and $(+-)_s$ categories. At any time t this pool contains $S_2=(\lambda_2+\lambda_3)(t-T_2)$ individuals in $(++)_s$. Therefore for $T_2 < t < T_3$

$$P[X_t=i'] = (R_1+R_2+S_2) \cdot P'[X_t=i']$$

$$= \left\{[1-F_6-(\lambda_2+\lambda_3)t] + (\lambda_2+\lambda_3)(t-T_2)\right\} \cdot P'[X_t=i']$$

$$= \left\{1-[F_6+(\lambda_2+\lambda_3)T_2]\right\} \cdot P'[X_t=i']$$

At $t=T_3$

$$P[X_{T_3}=i'] = (R_1+R_2+S_2/b_{iT_3}) \cdot P'[X_{T_3}=i']$$

$$= \left\{1-[F_6+(\lambda_2+\lambda_3)T_3] + (\lambda_2+\lambda_3)(T_3-T_2)/b_{iT_3}\right\} \cdot P'[X_{T_3}=i']$$

$$= [1-F_6-(\lambda_2+\lambda_3)(T_2-(T_3-T_2)/b_{iT_3})] \cdot P'[X_{T_3}=i']$$

Generalization of this expression gives the contribution of the first group to the expected total probability of detection as

$$\sum_{t=1}^{T_1-1} P'[X_t=i'] + \left\{1-[F_6+(\lambda_2+\lambda_3)T_1](1-\frac{1}{b_{iT_1}})\right\}\cdot P'[X_{T_1}=i'] +$$

$$\sum_{k=1}^{M}\left\{[1-F_6-(\lambda_2+\lambda_3)T_k]\left(\sum_{t=T_k+1}^{T_{k+1}-1} P'[X_t=i']\right) + Z_{T_{k+1}}\cdot\right.$$

$$\left.[1-F_6-(\lambda_2+\lambda_3)\left(T_{k+1}-\frac{T_{k+1}-T_k}{b_{iT_{k+1}}}\right)\cdot P'[X_{T_{k+1}}=i']\right\}$$

For the second group--those whose onset time is somewhere between the first and the second screening examination--the analysis is very similar:

For $T_1 < t < T_2$

$$P[X_t=i'] = (R_1+R_2+R_3)P'[X_t=i'] = P'[X_t=i']$$

At $t=T_2$

$$P[X_{T_2}=i'] = (R_1+R_2+R_3/b_{iT_2})\cdot P'[X_{T_2}=i']$$

$$= \left\{[1-F_6-(\lambda_2+\lambda_3)(T_2-T_1)] + \frac{F_6+(\lambda_2+\lambda_3)(T_2-T_1)}{b_{iT_2}}\right\}\cdot P'[X_{T_2}=i']$$

$$= \left\{1-[F_6+(\lambda_2+\lambda_3)(T_2-T_1)]\ (1-\frac{1}{b_{iT_2}})\right\}\cdot P'[X_{T_2}=i']$$

For $T_2 < t < T_3$

$$P[X_t=i'] = (R_1+R_2+S_2)\cdot P'[X_t=i']$$

$$= \Big\{ [1-F_5-F_6-(\lambda_1+\lambda_2)(t-T_1)]+[F_5+(\lambda_1-\lambda_3)(t-T_1) +$$

$$[(\lambda_2+\lambda_3)(t-T_2)]\Big\} \cdot P'[X_t=i']$$

$$= [1-F_6-(\lambda_2+\lambda_3)(T_2-T_1)] \cdot P'[X_t=i']$$

At $t=T_3$

$$P[X_{T_3}=i'] = (R_1+R_2+S_2/b_{iT_3}) \cdot P[X_{T_3}=i']$$

$$= \left\{ 1-F_6-(\lambda_2+\lambda_3)(T_3-T_1) + \frac{(\lambda_2+\lambda_3)(T_3-T_2)}{b_{iT_3}} \right\} \cdot P'[X_{T_3}=i']$$

In general, the contribution of the second group to the expected total probability of detection is

$$\sum_{t=T_1+1}^{T_2-1} P'[X_t=i'] + \left\{ 1-[F_6+(\lambda_2+\lambda_3)(T_2-T_1)(1-\frac{1}{b_{iT_2}})] \cdot P'[X_{T_2}=i'] + \right.$$

$$\sum_{k=2}^{M} \Big\{ [1-F_6-(\lambda_2+\lambda_3)(T_k-T_1)\Big(\sum_{t=T_k+1}^{T_{k+1}-1} P'[X_t=i'] \Big) + Z_{T_{k+1}} \cdot$$

$$\left. [1-F_6-(\lambda_2+\lambda_3) \left(T_{k+1}-T_1-\frac{T_{k+1}-T_k}{b'_{iT_k}} \right] \right] \Big\} \cdot P'[X_{T_{k+1}}=i']$$

The expression for the expected total treatment and death cost consists of the contribution of M+1 groups and is as follows

$$\sum_{j=1}^{M+1} W_j \left\{ \sum_{t=T_{j-1}+1}^{T_j-1} C_{ti'} \cdot P'[X_t=i'] + Z_{T_j} \cdot [1-(F_6+(\lambda_2+\lambda_3)(T_j-T_{j-1})(1-\frac{1}{b_{iT_j}})] \right.$$

$$\left. \cdot C_{ti'} \cdot P'[X_{T_j}=i'] + \sum_{k=j}^{M} \Big\{ [1-F_6-(\lambda_2+\lambda_3)(T_k-T_{j-1})] \right.$$

$$\sum_{t=T_k+1}^{T_{k+1}-1} C_{ti'} \cdot P'[X_t=i'] + Z_{T_{k+1}} \cdot$$

$$\cdot \left\{ \left[1-F_6-(\lambda_2+\lambda_3)(T_{k+1}-T_{j-1} - \left(\frac{T_{k+1}-T_k}{b_{iT_k}} \right) \right] \cdot C_{T_{k+1},i'} \cdot P[X_{T_{k+1}}=i'] \right\}$$

To find cost associated with the test itself, the cost associated with screening the categories $(-)_d$; $(+-)_d$; and $(++)_d$ should be added. Define

C_s = Cost of a single primary test used for mass screen.

$C_{s'}$ = Cost of secondary tests used/per individual.

C_t = Treatment cost/per patient.

$C_{t'}$ = Treatment cost/per suspect--a $(-)_d$ individual who responds $(++)_s$.

I = Number of population/per one patient.

The screening cost associated with $(-)_d$ group is

$$I[1-(F_1+F_2)]C_s+I\cdot F_1(C_s+C_{s'})+I\cdot F_2(C_s+C_{s'}+C_{t'}) =$$

$$I\cdot[C_s+(F_1+F_2)\cdot C_{s'}+F_2\cdot C_{t'}]$$

At any time t, the screening cost associated with the $(+-)_d$ group is

$$P[X_t=0]\left\{[1-(F_3+F_4)]\cdot C_s+F_3\cdot(C_s+C_{s'})+F_4\cdot(C_s+C_{s'}+C_{t'})\right\} =$$

$$P[X_t=0]\cdot[C_s+(F_3+F_4)\cdot C_{s'}+F_4\cdot C_{t'}]$$

For the $(++)_d$ group at $t=T_1$, the screening cost is

$$P[X_{T_1}\neq 0]\left\{[1-(F_5+F_6)-(\lambda_1+\lambda_2)T_1]C_s+[F_5+(\lambda_1-\lambda_3)T_1]\cdot(C_s+C_{s'}) + [F_6+(\lambda_2+\lambda_3)T_1]\cdot(C_s+C_{s'})\right\} =$$

$$P[X_{T_1} \neq 0] \left\{C_s+[F_5+F_6+(\lambda_1+\lambda_2)T_1]C_{s'}\right\}$$

At $t=T_2$, the screening cost is

$$P[X_{T_2} \neq 0] \left\{[1-(F_5+F_6)-(\lambda_1+\lambda_2)T_2]C_s + [F_5+(\lambda_1-\lambda_3)T_2](C_s+C_{s'})+[(\lambda_2+\lambda_3)(T_2-T_1)] (C_s+C_{s'})\right\} =$$

$$P[X_{T_2} \neq 0] \left\{[1-F_6-(\lambda_2+\lambda_3)T_1]C_s+[F_5+(\lambda_1+\lambda_2)T_2-(\lambda_1+\lambda_3)T_1]C_{s'}\right\}$$

Therefore the general expression for the expected total cost is

$$M.I.[C_s+(F_1+F_2)C_{s'}+F_2\cdot C_{t'}] + [C_s+(F_3+F_4)C_{s'}+F_4\cdot C_{t'}] \sum_{k=1}^{M} P[X_{T_k}=0] +$$

$$F_6\cdot(C_s+C_{s'})\cdot P[X_{T_1} \neq 0] + \sum_{k=1}^{M} P[X_{T_k} \neq 0] \left\{[1-F_6-(\lambda_2+\lambda_3)T_{k-1}]C_s + [F_5+(\lambda_1+\lambda_2)T_k-(\lambda_2+\lambda_3)T_{k-1}]C_{s'}\right\}$$

It should be mentioned that usually the first term is the only significant term. The expected total cost is composed of the test cost plus treatment and death cost.

In order to see the effect of using a double procedure to identify the diseased individuals in the optimal policy, a computer program was run in which the following probabilities, rates and costs were assumed to be known with their numerical values given as

$F_1=.002$ $F_2=.001$ $F_3=.04$ $F_4=.02$ $F_5=.1$ $F_6=.7$

$\lambda_1=.001$ $\lambda_2=.002$ $\lambda_3=.002$ rates/quarter

$C_s=.15$ $C_{s'}=4.00$ $C_{t'}=100.00$ dollars

The cost associated with different screening strategies was computed and the optimal policy was determined. To see the effect of variation in the probability of being in $(++)_s$, F_6 and associated rates were varied simultaneously over their range and results were compared with the case of constant F_6 and no rates of transition, which is the ordinary process studied before. This procedure was repeated for different disease parameters λ and b. The results are shown in Figures 20 and 21.

Taking $C_{s'}=C_{t'}=0$ is the same as assuming there is no extra charge for the secondary examination which is the same as ordinary system of screening seen before. Figure 20 shows that under the same conditions, any nonzero rate of transition to $(++)_s$ reduces the expected total cost of screening but has no effect on the optimal policy.

In some situations it might be a good idea for those individuals in $(+-)_s$ to be screened more often--they might be considered as high risk individuals--in which case a different analysis is needed.

Assuming that they are screened once more between each two scheduled screenings, the expression for $P[X_t=i']$ is

$$\sum_{j=1}^{M+1} W_j \left\{ \sum_{t=T_{j-1}+1}^{T_j-1} P'[X_t=i'] + Z_{T_j}\left[1-(F_6+(\lambda_2+\lambda_3)(T_j-T_{j-1})(1-\frac{1}{b_{iT_j}})\right]\cdot \right.$$

$$\left. \cdot P'[X_{T_j}=i'] \right\} + \sum_{k=j}^{M} \left\{ [1-F_6-(\lambda_2+\lambda_3)(T_k-T_{j-1})] \left(\sum_{t=T_k+1}^{T_{k'}-1} P'[X_t=i'] \right) + \right.$$

$$+Z_{T_{k'}}\left[1-F_6-(\lambda_2+\lambda_3)(T_{k'}-T_{j-1})+\frac{(T_{k'}-T_{k-1})\lambda_3}{b_{iT_{k'}}}\right]\cdot P'[X_{T_{k'}}=i']$$

$$\left. +[1-F_6-\lambda_2(T_k-T_{j-1})-\lambda_3(T_{k'}-T_{j-1}) \left(\sum_{t=T_{k'}+1}^{T_{k+1}-1} P'[X_t=i'] \right) + \right.$$

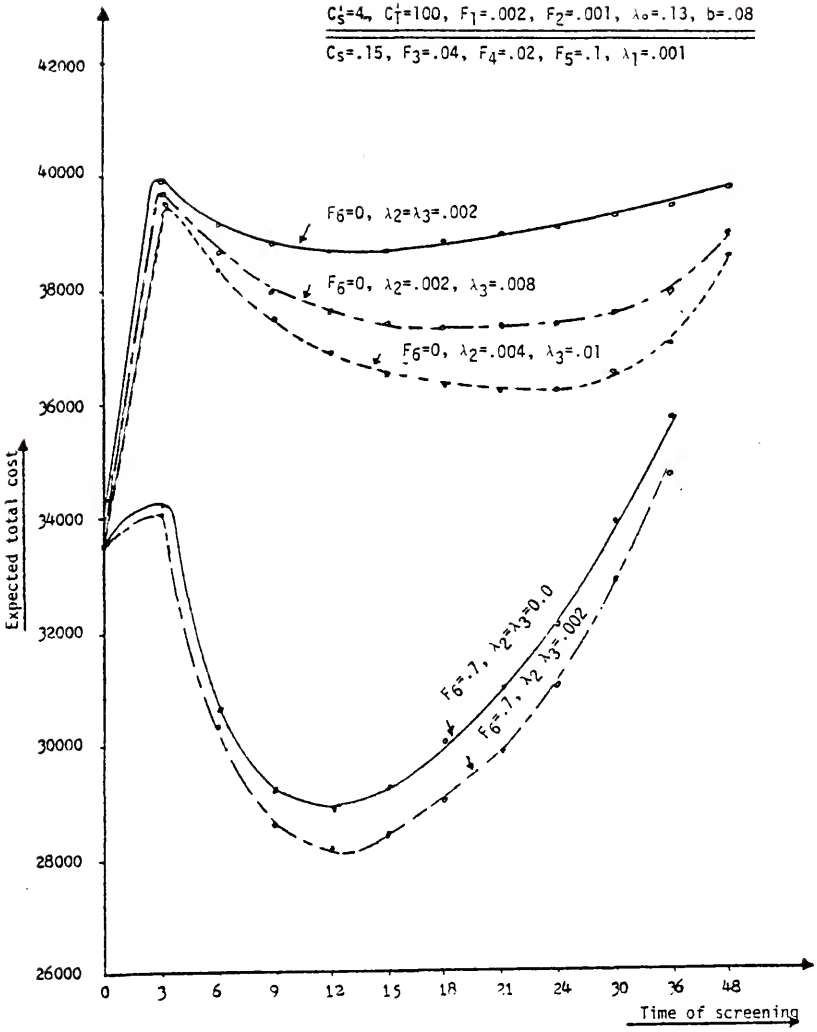


Figure 20: Comparison between different rates and probabilities of being in$(++)_S$.

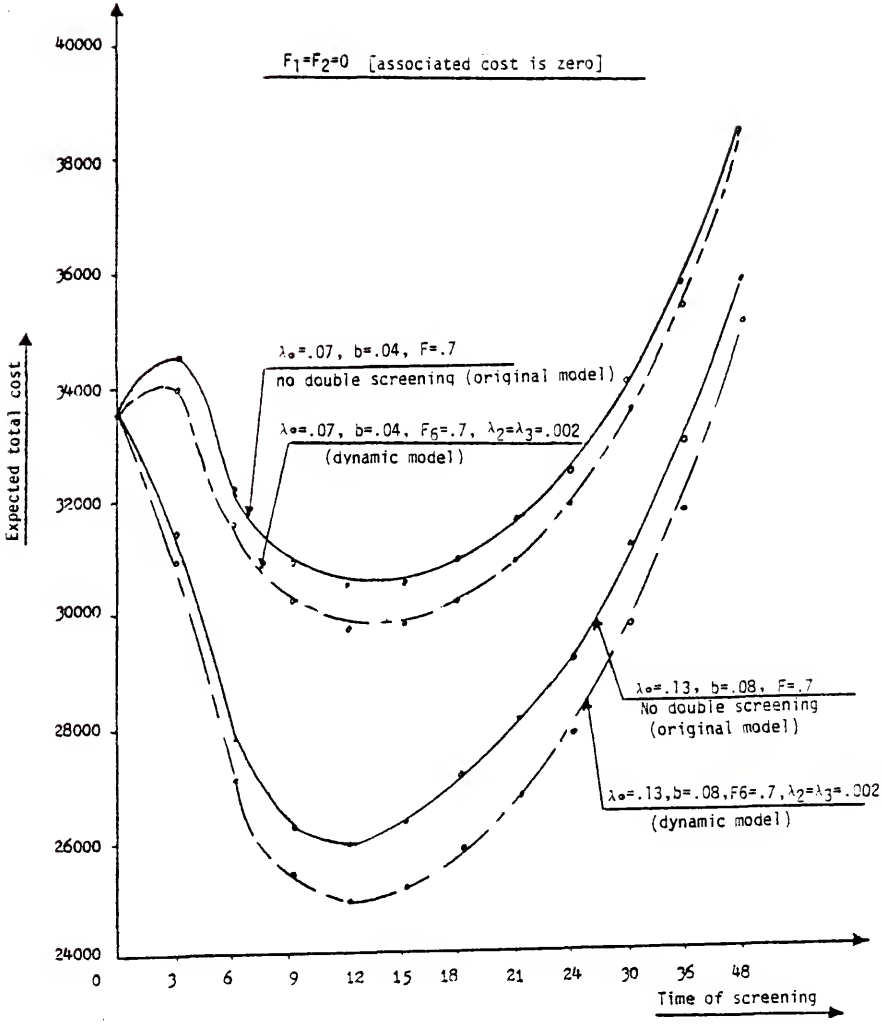


Figure 21: Comparison with ordinary system.

$$+ Z_{T_{k+1}} \cdot \left[1-F_6-(\lambda_2+\lambda_3)(T_{k+1}-T_{j-1}- \frac{(\lambda_2+\frac{\lambda_3}{2})(T_{k+1}-T_k)}{b_{iT_{k+1}}}\right] \cdot$$

$$\cdot P'[X_{T_{k+1}}=i']\Big\}\Big\}$$

where $T_{k'} = \frac{T_k+T_{k+1}}{2}$.

5.4 Transient Problem

Up to now, in any analysis, a population was considered which had been under a scheduled screening program for a long period of time--long enough to make sure that every individual has had the opportunity of participating in all examinations. In such a case, it was mentioned that objective function consists of contributions from M+1 groups and every individual has participated in all M examinations but those who were in group i had M+1-i times the chance of being detected. As Albert, et al. mention [2,3,66] when a screening strategy is established the process goes through a transient period during which steady state analysis is not valid. To extend the analysis to such a case, the following argument is employed.

Assume

1) M screenings are done at ages $\hat{T}_1$, $\hat{T}_2=\hat{T}_1+dT$, $\hat{T}_3=\hat{T}_2+dT,\ldots,\hat{T}_M=\hat{T}_{M-1}+dT$ (extension to the case of unequal screening interval is not difficult).

2) As soon as an individual reaches ages $\hat{T}_1$, $\hat{T}_2,\ldots,\hat{T}_M$, he is screened.

In this analysis, the time from the beginning of the screening program, denoted by T, plays a role so that expected screening cost at any time T

is a function of T. Figure 22 is a sketch which represents the situation. In this plot the horizontal lines are representative of ages at screening and the 45° lines represent growth of an individual. It is seen that as T increases to T+dT, age t increases to t+dT too.

There are M+1 intervals of interest. The i-th interval is defined to be all T's such that $\hat{T}_i-\hat{T}_1 < T \leq \hat{T}_{i+1}-\hat{T}_1$ or equivalently $\hat{T}_i < T+\hat{T}_1 \leq \hat{T}_{i+1}$. In each interval there is a certain pattern that changes from interval to interval. For instance, in the first interval the pattern is (0,1), which means

1) At any time $0 < T < \hat{T}_2-\hat{T}_1$, an individual of age $t < \hat{T}_1$ has had no tests in his lifetime,

2) An individual of age $\hat{T}_1 \leq t \leq \hat{T}_1+T$ has had one test in his lifetime,

3) An individual of age $\hat{T}_1+T < t < \hat{T}_2$ has had no test in his lifetime,

4) An individual of age $\hat{T}_2 \leq t \leq \hat{T}_2+T$ has had one test in his lifetime.

This is because an individual of age $\hat{T}_1 \leq t \leq \hat{T}_1+T$ has been $\hat{T}_1$ years old after the beginning of the screening program and therefore has participated in the first examination at age $\hat{T}_1$, whereas an individual of age $\hat{T}_1+T < t < \hat{T}_2$ has been older than $\hat{T}_1$ and yet younger than $\hat{T}_2$ and therefore has had no chance of being screened.

In the second interval the pattern is (0,1,2), which means

1) At any time $\hat{T}_2-\hat{T}_1 < T \leq \hat{T}_3-\hat{T}_1$, an individual of age $t < \hat{T}_1$ has had no test, but

2) An individual of age $\hat{T}_1 \leq t < \hat{T}_2$ has had one test.

3) An individual of age $\hat{T}_2 \leq t \leq \hat{T}_2+T$ has had two tests.

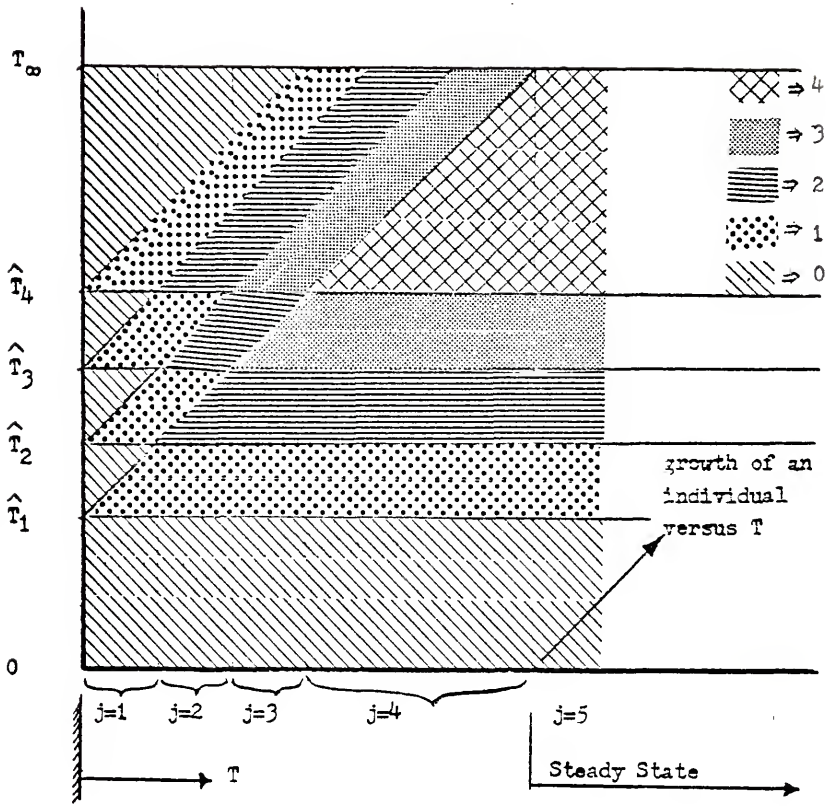


Figure 22: A sketch of the transient problem. T is the time measured from the start of the screening program.

4) An individual of age $\hat{T}_1+T < t < \hat{T}_3$ has had one test.

This argument is only for the number of tests given to an individual. To see the benefit of screening, it is essential to look at each group as a separate entity and find the contribution of all M+1 groups. Obviously for the first group--those who have the disease before the first screening time--the benefit is the same as the number of screens, as shown in Figure 22. For second, third, fourth and ... group the sketch is shown in Figure 23.

The derivation of the general expression is given in the Appendix. It is seen that objective function depends on T and changes from interval to interval.

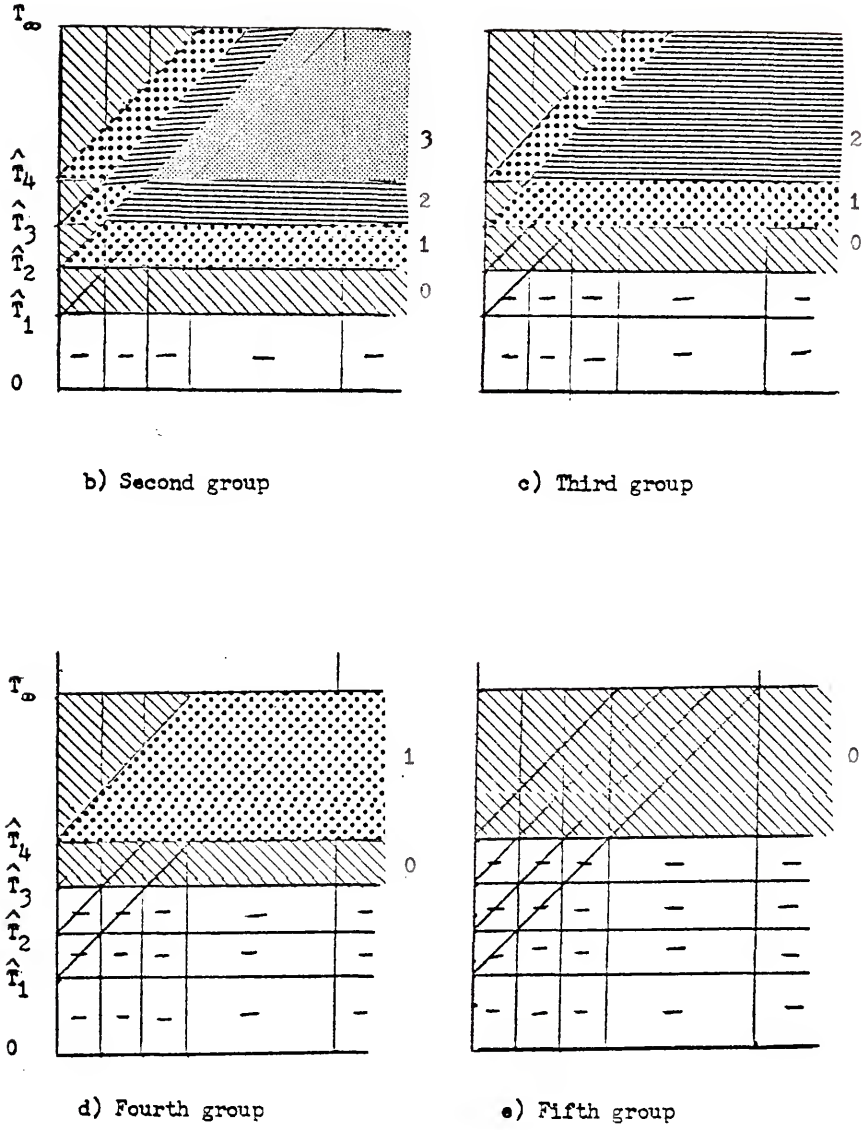
It is seen that steady state is reached for $T \geq t_\infty - \hat{T}_1$, because for any T in this interval every individual in the original population has been screened M+1 times and there is no further transient effect. This can be seen from the mathematical expression for the transient problem for the case j=M+1. In this case (j=M+1) the objective function reduces to

$$O.F.(M+1) = \sum_{i=1}^{M+1} W_i \cdot \left\{ \sum_{\theta=0}^{M} \left[\sum_{t=T_{(\theta+i-1)}+1}^{T_{(\theta+i)}} (1-F)^{\theta} \cdot A_t \right] \right\}$$

which is exactly the same as that derived for the steady state situation.

To see the effect of the transient time from the date that the screening program has started, a computer program was run in which time T was varied from zero to a point when steady state conditions were reached. Figures 24 and 25 are plots of the expected total cost of screening per one patient for the case of one and two screenings per lifetime.

It is seen that as time increases, expected total cost of screening increases continuously until it reaches its steady state value at



b) Second group

c) Third group

d) Fourth group

e) Fifth group

Figure 23: A sample problem for case $\begin{cases} M=4,\ t_{max}=9, \\ T_1=2,\ dT=1 \end{cases}$

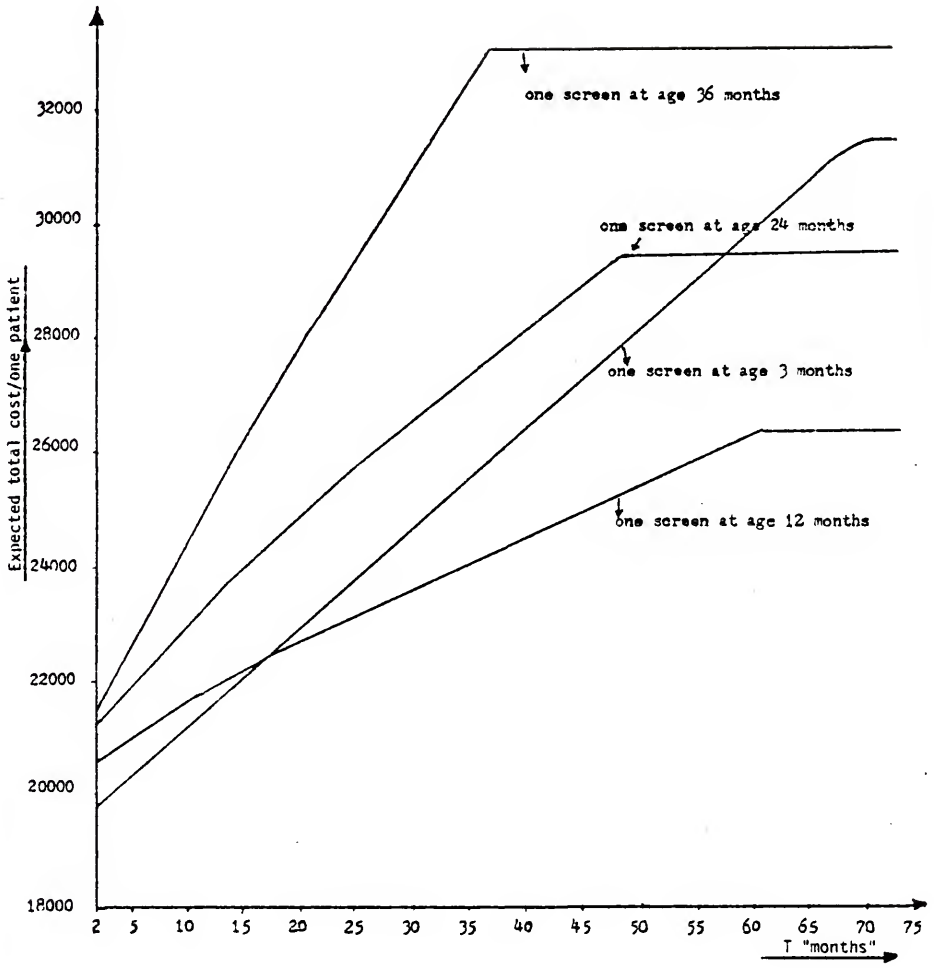


Figure 24: The expected total cost of screening per one patient as a function of time T for one screen per lifetime. T is the time measured from the date that the screening program was started.

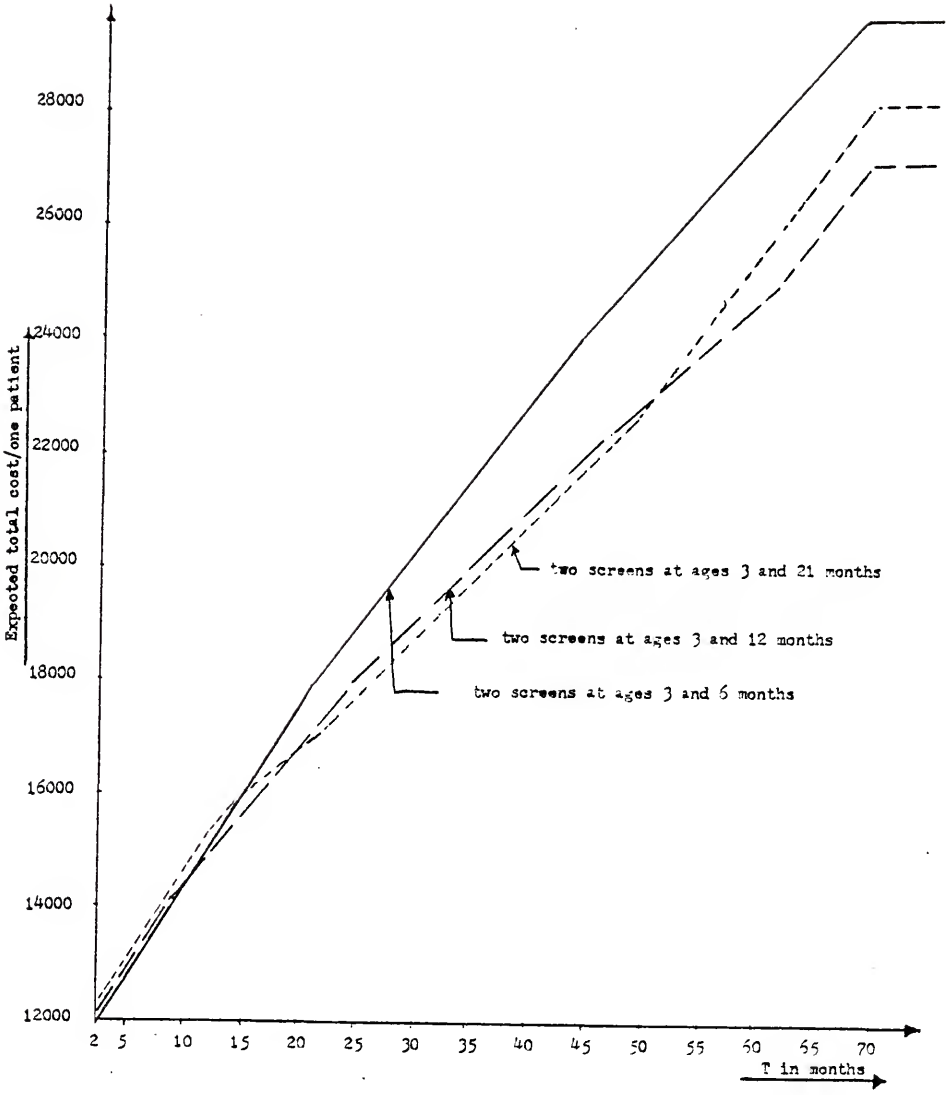


Figure 25: The expected total cost of screening per one patient as a function of T for two tests per lifetime , the first one of which is done at age 3 months.

$T=t_{max}-\hat{T}_1$. This is explained by the fact that at the beginning of the transient period there are more diseased individuals and therefore screening is more beneficial. As time increases, the prevalence decreases until at $T=t_{max}-\hat{T}_1$ it reaches its steady state value.

A very similar expression exists for the case of unequal screening intervals. The only difference in analysis is in the definition of the intervals. In the case of unequal screening intervals, it would be necessary to look at such intervals as: $(0,\hat{T}_2-\hat{T}_1)$, $(\hat{T}_2-\hat{T}_1, \hat{T}_3-\hat{T}_2)$, $(\hat{T}_3-\hat{T}_2, \hat{T}_3-\hat{T}_1)$, $(\hat{T}_3-\hat{T}_1, \hat{T}_4-\hat{T}_2)$, $(\hat{T}_4-\hat{T}_2, \hat{T}_4-\hat{T}_1)$,....

5.5 Application of the Model to the Case of Breast Cancer Using the Results of HIP Study

In order to validate any mathematical screening model, there is a need for a set of data on a progressive disease that can be used as input so that the model's output can be compared with that data to check for any discrepancy. In case of Neuroblastoma there are no such data. In fact in the case of most progressive diseases there is little that can be said about a complete set of data. As was mentioned in the literature review, it is not possible to gather complete data on the nature of any malignant disease such as cancer, because some detected diseased individuals would have to go untreated. Therefore it remains to base the analysis on certain assumptions on the behavior of the disease in the occult part. Having assumed the pattern of the behavior of the disease in the occult part, it would be easy to get an estimate of the parameters by varying them until the system's output matches the available data. Then under those specific assumptions, the values of the parameters are known. This type of analysis has been used by different investigators including Knox [55,56] in his

simulation model and Shwartz [88,89,90,91,92,93] in validation of his model through use of the HIP data for breast cancer.

In the case of our model, the HIP study which has gathered data on 62,000 women in two groups, a study group and a control group, can be employed for the breast cancer. The model needs some information on the natural probability of detection which has to be gathered while the system has been under no scheduled external screening program. Luckily HIP data on the control group are gathered in the manner that the model requires. Therefore the application of the model to the breast cancer using the HIP study is relatively simple. The procedure is to use data on the control group as that of the population under no scheduled screening program. Then by introducing a scheduled screening program similar to that of HIP the methodology is implemented to the control group and a set of effectiveness measures are recorded. These outputs are then compared with corresponding data on the study group and some conclusions are made.

The HIP screening schedule was to go through four annual screening examinations but not everybody in the study group participated in the program. Even some of those who participated in the initial examination did not participate all or part of the rest of the program. The actual screening pattern and percentage of women in each category is shown in Table 9, extracted from reference [29].

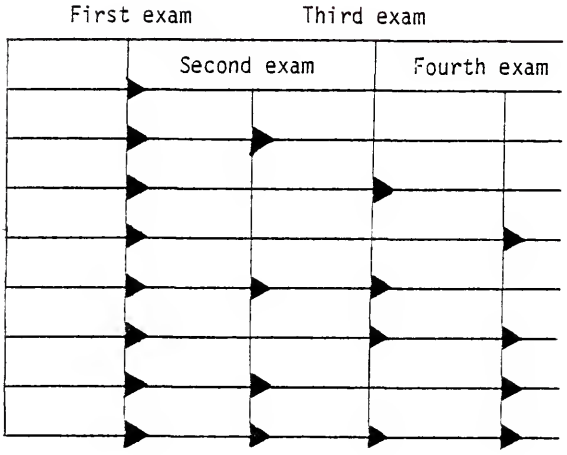
In order to simplify the procedure, all patterns with the same number of examinations were added up and therefore it was assumed that there are five patterns as follows:

1) No screen (%35)

2) One screen (% 8) assumed to be done at initial examination.

3) Two screens (% 8) assumed to be done at initial examination and one year after that.

Table 9: The percentage of individuals in each pattern of screening.



No screen ratio: .3495
One screen ratio: .0852
Two screens ratio: .0468
Two screens ratio: .0136
Two screens ratio: .0136
Three screens ratio: .0553
Three screens ratio: .0214
Three screens ratio: .0261
Four screens ratio: .3864

Table 10: Data on $P'[X_t=D]$

Age at diagnosis	Estimated population at risk	No. of breast cancers	Incidence
40-44	6955	7	$1.01(10)^{-3}$
45-49	7472	9	$1.2\ (10)^{-3}$
50-54	6962	9	$1.29(10)^{-3}$
55-59	5999	12	$2.00(10)^{-3}$
50-64	3797	8	$2.11(10)^{-3}$
Total		45	

4) Three screens (%11) assumed to be done in first three consecutive years.

5) Four screens (%38) assumed to be done in all years.

Of course in reality a woman who participated in two annual examinations might have had an interval of up to three years between the two tests. In the model it is assumed that the interval between two successive exams is fixed at one year. This might reduce the efficiency of screening but the reduction should not be significant.

In the HIP study only women in the age group 40-64 were selected. This type of selection makes it possible for a case to go into screening at ages 40, 41, 42 and 43 whereas another case might go into screening at ages 61, 62, 63 and 64. But the model used assumes that everybody in the population is screened at specific ages $\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_M$. One way to handle this situation is to introduce different age groups, say 40-44, 45-49,... and to find the expression for each age group. The weighted sum of all the age groups reveals the quantity of interest. Another way is to assume that every individual has an average age 52=(40+64)/2 and is screened at ages 52, 53, 54 and 55. This procedure is more of an approximation but needs less calculation and is, therefore, preferred. In our analysis both methods were employed and some sensitivity analysis was done on the results.

According to the data, it is necessary to consider cases that did not participate in all the scheduled examinations. Therefore the following analysis was employed. It considers different screening patterns:

Define $R(\ell,k)$ to be the ratio of population who have completed $(\ell-k)$ examinations up to ℓ-th interval (the i-th interval is the time

between the (i-1)st and the i-th exam).

Then for $t < \hat{T}_1$: $P[X_t=D] = W_1\left\{R(1,1)\cdot P'[X_t=D]\right\}$

where $R(1,1)=1$.

At $t=\hat{T}_1$: $P[X_{T_1}=D]=W_1\left\{R(2,1)(1-F+\frac{F}{b_{iT_1}})+R(2,2)\right\}P'[X_{T_1}=D]$

where $R(2,1)$ = percent of participants in the first exam,

and $R(2,2)$ = percent of those who never participated in the screening program.

For $\hat{T}_1 < t < \hat{T}_2$: $P[X_t=D] = W_1\left\{R(2,1)(1-F)+R(2,2)\right\} P'[X_t=D]$

$$+ W_2\left\{R(2,1)+R(2,2)\right\} P'[X_t=D]$$

where W_1 and W_2 are the weights associated with first and second groups.

It is seen that for the first group only a percent (1-F) of the participants remain undetected after the first exam, whereas for the second group, irrespective of whether or not they participated in the first exam, they are detected under their natural probability of detection--simply because they were not yet sick at the time of the first exam. A similar analysis for other time intervals is valid:

i.e. at $t=\hat{T}_2$: $P[X_{T_2}=D] = \left\{W_1[R(3,1)(1-F)(1-F+\frac{F}{b_{iT_2}}) + R(3,2)(1-F)+R(3,3)] + W_2[R(3,1)(1-F+\frac{F}{b_{iT_2}})+R(3,2)+R(3,3)]\right\}P'[X_{T_2}=D]$

At $t=\hat{T}_3$: $P[X_{T_3}=D] = \left\{W_1[R(3,1)(1-F)^2+R(3,2)(1-F)+R(3,3) + W_2[R(3,1)(1-F)+R(3,2)+R(3,3)] + W_3[R(3,1)+R(3,2)+R(3,3)]\right\} P'[X_{T_3}=D]$

For $\hat{T}_3 < t < \hat{T}_4$: $P[X_t=D] = \Big\{ W_1[R(4,1)(1-F)^3+R(4,2)(1-F)^2+R(4,3)(1-F)+R(4,4)]$

$+ W_2[R(4,1)(1-F)^2+R(4,2)(1-F)+R(4,3)+R(4,4)]$

$+ W_3[R(4,1)(1-F)+R(4,2)+R(4,3)+R(4,4)]$

$+ W_4[R(4,1)+R(4,2)+R(4,3)+R(4,4)] \Big\} P'[X_t=D]$

and so on.

Therefore the general expression would be

$$P[X_t=D] = \sum_{j=1}^{m_t} W_j \left\{ \sum_{k=1}^{l_t} R(1,k) \cdot (1-F)^{S(j,k,t)} \underbrace{\cdot (1-F+\frac{F}{b_{it}})^{Z_t}}_{\text{if } k=1} \right\} \cdot P'[X_t=D]$$

where : $m_t = \sum_{i=1}^{t-1} Z_i + 1$

$$l_t = \sum_{i=1}^{t} Z_i + 1 = m_t + Z_t$$

and $S(j,k,t) = \text{Max} \begin{cases} \left\{ \sum_{i=\hat{T}_{j-1}+1}^{t-1} Z_i - k + 1 \ , \ 0 \right\} & \text{if } k=1 \\ \left\{ \sum_{i=\hat{T}_{j-1}+1}^{t} Z_i - k + 1 \ , \ 0 \right\} & \text{if } k \neq 1 \end{cases}$

It should be mentioned that to get an expression for cost effectiveness measurements, the expression for $P[X_t=D]$ is modified by the corresponding cost factors and summed over all time intervals.

The expression for $P[X_t=D]$ is a generalization of the original expression for expected total cost under the condition that some individuals skip some of the scheduled screening tests. If every individual participates in every examination, then $R(i,1)=1$ and $R(i,j)=0 \ \forall \ j \neq 1$.

Under these circumstances the general expression reduces to

$$P[X_t=0] = \sum_{j=1}^{m_t} W_j \, (1-F)^{\sum_{t=\hat{T}_j+1}^{t-1} Z_t} \left(1-F+\frac{F}{b_{it}}\right)^{Z_t} \cdot P'[X_t=0]$$

which is the same as the original expression.

The general expression could be employed to the case of any type of screening program in which some individuals do not participate in all the scheduled examinations regularly. In particular data [29] show that in the case of HIP study the $R(\ell,k)$ matrix looks like the following:

$$R(\ell,k) = \begin{bmatrix} R(1,1)=1.0 & & & & \\ R(2,1)=.65 & R(2,2)=.35 & & & \\ R(3,1)=.57 & R(3,2)=.08 & R(3,3)=.35 & & \\ R(4,1)=.49 & R(4,2)=.08 & R(4,3)=.08 & R(4,4)=.35 & \\ R(5,1)=.39 & R(5,2)=.10 & R(5,3)=.08 & R(5,4)=.08 & R(5,5)=.35 \end{bmatrix}$$

It is seen that $R(\ell,1)$ is always divided into $R(\ell+1,1)$ and $R(\ell+2,2)$, i.e.

$$R(\ell,1)=R(\ell+1,1)+R(\ell+2,2)$$

and $$R(\ell+1,k+1)=R(\ell,k) \quad \forall \quad \ell > k > 1$$

which makes it very easy to construct the $R(\ell,k)$ matrix--start from last row.

One essential input of the model is the natural probability of being in the detected state, $P'[X_t=0]$, which is a characteristic of the disease under no scheduled external interference. In the case of HIP study data in Zelen and Feinleib [116] or [2] could be used. The difficulty in the use of this data is in its nonspecificity. Data as

seen in Table 10 are presented in a very broad range. Therefore in order to employ it, some assumptions on the form of behavior of $P'[X_t=D]$ versus age has to be made.

Since data are only for ages 40-64, it is assumed that a woman is naturally detected only in this age group. In fact, the process is truncated on both sides. As a result, a probability distribution for $P'[X_t=D]$ is found by calculating $P'[X_t=D]$ for each interval and fitting a least square line through the points.

Total incidence = $(1.01+\ldots+2.11)\cdot(10)^{-3}=7.61\cdot(10)^{-3}$

$P'[X_t=D]$ for age group 40-44 $=1.01(10)^{-3}/7.61(10)^{-3}=.133$

$P'[X_t=D]$ for age group 45-49 $=1.20(10)^{-3}/7.61(10)^{-3}=.158$

$P'[X_t=D]$ for age group 50-54 $=1.29(10)^{-3}/7.61(10)^{-3}=.169$

$P'[X_t=D]$ for age group 55-59 $=2.00(10)^{-3}/7.61(10)^{-3}=.263$

$P'[X_t=D]$ for age group 60-64 $=2.11(10)^{-3}/7.61(10)^{-3}=.277$

If the system is observed every three months, then the equation of the polynomial of degree two and the straight line would be

$$P'[X_t=D] = (3553.68-35.93t+.1525\cdot t^2)/280000$$

$$P'[X_t=D] = (.89375\ t+54.1447)/10000$$

where t is age in quarters from 40 years on. (For instance t for a 45 years old woman would be $t=4(45-39\frac{3}{4})=21$. In other words, $T=39\frac{3}{4}+\frac{t}{4}$ where T is the actual age in years.)

If the system is observed every four months, then the equation of the straight line would be

$$P'[X_t=D] = (1.5889\ t+72.19293)/10000$$

where t is age in four months from 40 years on.

To find an estimate of the true positive rate of screening, data on the study group of HIP were employed. In the process of examination, each individual has been screened by two different and independent methods--mammography and clinical palpation. Any positive result would have brought the patient into attention. Assuming that the true positive rate of those two methods is F_1 and F_2 and letting F_{12} be the probability that a patient gives positive responses to both examinations reveals : $F = F_1 + F_2 - F_{12}$

HIP data [87] show that for $40 < t < 49$ $F=19.4+61.3=80.7$

for $50 < t < 59$ $F=41.5+40.0=81.5$

for $60 > t$ $F=30.6+38.9=69.5$

In the calculation, F was chosen to be a constant F=0.8. Later, F was varied to see the effect of reducing the test sensitivity on the screening effectiveness measures.

The incidence rate--number of population which gets the disease in the age range 40-64--was established using data on HIP [2,3] for the control group.

Incidence = 1.6 per thousand per year = 8/1000 (5 years). Other estimates are available. For instance Shwartz [88,89,90,91,92,93] gives

Incidence = (102.8+156+168.4+188.3+221.5)/100000=8.37/1000 (5 years)

As an estimate it was assumed that in five years (the time in which the HIP data were collected) the incidence of breast cancer in the women of age range 40-64 is one in every 104 individuals. This value was varied to see its effect on the estimation of the effectiveness measurements.

A common problem in the analysis of all the progressive diseases is the lack of understanding of the behavior of the disease in the occult

part. In particular there is no way to find the distribution of the onset. Therefore it remains to assume a reasonable distribution for onset. Reasonable here means it should be so that the cumulative onset at a time t is always greater than the cumulative probability of natural detection at t.

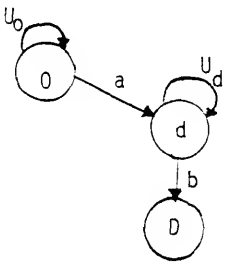
In the model, it is necessary to have an estimate of the onset distribution and b_t, which is the natural probability of transition from the occult state to the detected state. They are not independent quantities. In fact as will be shown, determination of one of them fixes the value of the other. The argument is based on the structure of the model.

The disease has the following structure:

where 0 = Healthy state

d = Occult state

D = Detected state



At any time t : $U_{0t}+a_t=1$

$U_{dt}+b_t=1$

But $P[X_t=D]=P[X_t=D|X_0=0]\ P[X_0=0]+P[X_t=D|X_0=d]\ P[X_0=d]$

Define δ_o to be $P[X_0=0]$ which is probability of being healthy at age zero. Employing observations 1-4, it is easy to show that

$$P[X_t=D] = \left\{ \sum_{j=1}^{t-1} \left(\prod_{i=1}^{j-1} U_{0i} \right) a_j \cdot \left(\prod_{k=j+1}^{t-1} U_{dk} \right) \delta_o + \left(\prod_{j=1}^{t-1} U_{dj} \right) (1-\delta_o) \right\} \cdot b_t$$

But

$$U_{ot}=1-a_t$$

$$U_{dt}=1-b_t$$

Therefore

$$P[X_t=D] = \left\{ \sum_{j=1}^{t-1} \left(\prod_{t=1}^{j-1} (1-a_i) \right) \cdot a_j \cdot \left(\prod_{k=j+1}^{t-1} (1-b_k) \right) \delta_o + \left(\prod_{j=1}^{t-1} (1-b_j) \right) (1-\delta_o) \right\} b_t$$

This is a system of equations in which the a's and b's are unknown and $P'[X_t=D]$ is a known quantity, i.e.,

For t=1 $\quad P'[X_1=D]=(1-\delta_o)\cdot b_1=C_1\cdot b_1 \quad \Rightarrow \quad b_1=P'[X_1=D]/C_1$

For t=2 $\quad P'[X_2=D]=[a_1\delta_o+(1-b_1)(1-\delta_o)]\cdot b_2=C_2 b_2 \Rightarrow \quad b_2=P'[X_2=D]/C_2$

For t=3 $\quad P'[X_3=D]=\{[a_1(1-b_2)+(1-a_1)a_2]\delta_o+(1-b_1)(1-b_2)(1-\delta_o)\}\, b_3$

$$\Rightarrow \quad b_3=P'[X_3=D]/C_3$$

$\vdots$

Therefore knowing $P[X_t=D]$, the above analysis can be employed to generate the values for the probability of transition from d to D. In the case of breast cancer it remains to find an appropriate distribution for the onset time.

In order to see the implications of the model to the case of breast cancer, two computer programs were written:

PROGRAM 1 : Every individual has an average age of 52 at the beginning of the program and her age at scheduled annual examinations is 52, 53, 54 and 55.

PROGRAM 2 : The population is divided into the following five groups

40-44, 45-49, 50-54, 55-59, and 60-64.

5.1.1 Program 1

In the primary analysis, a set of legitimate onset distributions were used and the output parameters of interest were recorded. These parameters are

1) The probability of natural detection at any age t--known from data on control group.

2) The probability of detection (under screening program) at any age t.

3) The cumulative probability of natural detection.

4) The cumulative probability of detection (under screening program).

5) The number of individuals detected at any time t under natural detection.

6) The number of individuals detected at any time t under screening program.

7) The number of individuals detected up to time t under natural detection.

8) The number of individuals detected up to time t under screening program.

Table 11 summarizes the computer output for different values of onset distribution. In this table

1) Onset distribution is given by five numbers, the i-th number representing the cumulative onset rate in the i-th interval. For instance $(a_1, a_2, a_3, a_4, a_5)$ means a_i percent get the disease somewhere between (i-1)st and i-th test.

2) Number of women detected due to the initial test = N_1.

3) Number of women detected due to annual exams = N_2.

4) Total number of women detected in the control group = N_c.

Table 11: Comparison of model's output with HIP.

Model	Onset Distribution					N_1	N_2	N_c	N_s
1	.57	.4	.02	.01	.0	53.17	69.79	283.17	307.29
2	.5	.2	.2	.1	.0	45.58	76.91	283.17	320.17
3	.6	.15	.15	.1	.0	56.43	65.11	283.17	314.5
4	.45	.4	.1	.05	.0	40.15	82.04	283.17	315.13
5	.45	.35	.15	.05	.0	40.15	82.85	283.17	317.45
6	.5	.4	.07	.02	.01	45.58	76.15	283.17	311.66
7	.53	.32	.07	.04	.02	48.83	69.99	283.17	310.81
8	.45	.45	.08	.02	.0	40.15	81.91	283.17	312.34
9	.6	.3	.08	.02	.0	56.43	67.41	283.17	309.60
10	.6	.35	.04	.01	.0	56.43	67.19	283.17	307.37
11	.54	.4	.03	.02	.01	49.92	71.64	283.17	309.11
12	.56	.39	.04	.01	.0	52.09	71.07	283.17	308.29
13	.4	.4	.18	.01	.005	34.73	87.90	283.17	317.87
14	.44	.24	.21	.11	.0	40.26	57.65	283.17	297.90
HIP	———					55	77	284	296

5) Total number of women detected in the study group = N_s.

It is seen that within some range all the cases generated values which are comparable with the corresponding values of HIP study. To see this better, consider case 1 and find some parameters of interest, using this particular case study, and compare them with corresponding values of HIP.

1) Data show detection due to initial exam = 55

The model shows detection due to initial exam = 53.17 (% error=3)

2) Data show detection due to annual exam = 77

The model shows detection due to annual exam = 69.79 (% error=9)

3) Data show 296 were diagnosed among 31000 study group

The model shows 307.22 were diagnosed among 31000 study group (% error=3)

4) Data show 284 were diagnosed among 31000 control group

The model shows 283.17 were diagnosed among 31000 control group (% error=0)

5) Data show detected people in initial exam had prevalence 2.72/1000=55/20200

The model shows detected people in initial exam had prevalence 2.64/1000=53.17/20200 (% error=2)

6) Data show detected women in rescreen exam = 1.51/1000 person/year

The model shows detected women in rescreen exam = 1.55/1000 person/year (% error=2)

7) Data show incidence rate among control group = 1.86/1000

The model shows incidence rate among control group = 1.83/1000 (% error=1)

8) Data show the number of study women refused test who got cancer = 73. Unfortunately, the model does not consider biasness in acceptance or rejection of program. Therefore a very rough estimate would be to assume

that they had a detection probability similar to control group (this is not true [31]). Therefore the number of study women refused test and contracted cancer = 99. This gives an average incidence rate of 1.83/1000 for study women refusing test, whereas data show average incidence rate of 1.37/1000. The discrepancy is in the biasness in selection process which is not incorporated in the model.

9) Data give the average age of the screened group at diagnosis = 54.4
The average age of the study group refusing test = 53.3
The average age of the control group = 54.5

Unfortunately the assumptions of Program 1 make it impossible to compute this quantity. Therefore it was decided to modify the program to increase its ability to compute average age at detection--the new modified program is called Program 2.

5.5.2 Program 2

In this program the total contribution to any effectiveness measure is the weighted sum of the contribution from each category where the weighting factor is the percentage of women in each age group determined from HIP data [89]. Data show that the population age is not uniform and has the distribution which is shown below:

Age	HIP sample size	Weight
40-44	4325	%21
45-49	4770	%24
50-54	4648	%23
55-59	3800	%19
60-64	2668	%13
	Total: 20200	% 100

This says 21% of the women who were in the study group were assumed to be in the 40-44 age group.

The incidence rate for all age groups was taken to be the same as that of Program 1.

Moreover in this program the same quantities as that of Program 1 were computed for all age groups. Then using the weighted sum of quantities the corresponding values for the general population was calculated.

Table 12 summarizes the computer output for different onset distributions. Table 13 gives a comparison between the model's output and that of the study group in the HIP study for case number 1.

To compare the model's ability to predict a process with that of other investigators, a comparison is done with the Shwartz model (he has presented two models which differ in their distribution of tumor growth). Table 14 shows the HIP findings along with our model's and the Shwartz predictions.

Table 12: Summarization of the computer output for different onset distributions.

HIP finding	55	77	284	296

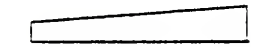
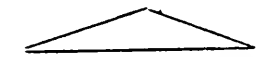
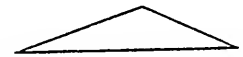
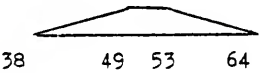
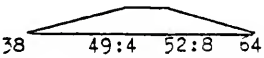
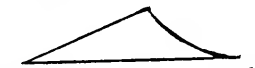
Case	Onset distribution	N_1	N_2	N_c	N_s
1		55.3	76.8	285	290.3
2		55.3	80.3	285	287.1
3		55.3	76.0	285	282.5
4		55.3	76.6	285	283.0
5		55.3	77.7	285	284.1
6		55.3	76.0	285	282.5
7		55.3	78.6	285	282.8

Table 13: Comparison between the model's output and data on HIP study (case no. 1)

No.	Subject of comparison	Data	Model	% error
1	Detection due to initial exam	55	55.3	0.5
2	Detection due to annual exam	77	76.8	0.2
3	Women diagnosed (control group)	284	285	0.3
4	Women diagnosed (study group)	296	290.4	1.8
5	Prevalence of the detected women in initial exam	$\frac{2.72}{1000}$	$\frac{2.73}{1000}$	0.3
6	Incidence of women in rescreen exam (person-year)	$\frac{1.51}{1000}$	$\frac{1.506}{1000}$	0.2
7	Incidence rate among control group	$\frac{1.86}{1000}$	$\frac{1.832}{1000}$	1.0
8	Number of women diagnosed in the study group			
	Initial exam	55	55.3	
	First annual	33	32.2	
	Second annual	21	25.8	
	Third annual	23	18.8	
9	Average age of the screened group at diagnosis	54.5	54.45	0.0
10	Average age of those detected by screen in the study group	54.9	54.1	1.4
11	Average age of those surfaced between screens in study group	54.0	54.75	1.3

Table 14: Comparison between our model and Shwartz model on predicting the HIP findings.

Detection rates per 1000	HIP rate	Our model	Shwartz model #1	Shwartz model #2
Screened group Detection on initial exam	2.72	2.73	2.50	2.49
Detection on rescreen	1.51	1.506	1.58	1.61
Detection not due to screen	.92	.88	1.0	.97
Control group	1.86	1.832	1.75	1.75

CHAPTER 6

DETERMINATION OF AN OPTIMAL POLICY

In a screening program the cost-effectiveness objective function is a function of the decision variables Z_t, which can take on integer values (0,1). This reduces the optimization problem to an integer program, but due to nonlinearity of the objective function, it is not a linear integer program and there is no single known method for its solution. In most searches for determination of the optimal objective function in the literature and in this research up to now, the determination of an optimal policy has been postponed and the search has been directed at detection of a "good" policy within a set of conventional policies doing screening every DT time units, DT=1,2,... There are some indications that an optimal screening policy need not always be in the set of conventional solutions. In this chapter an attempt is made to solve the optimization problem and offer an efficient methodology, if possible, to determine the "best" policy. In section 6.1 a branch and bound approach is employed for which 2 different lower bound approximation methods are used. Section 6.2 presents a heuristic approach to the determination of a "good" solution.

6.1 Branch and Bound Method: Search for an Optimal Solution

The expression for the expected total cost of screening is

$$OF = \sum_{j=1}^{M+1} \left\{ \sum_{t=T_{j-1}+1}^{t\infty} W(j,t)\cdot(1-F)^{\sum_{k=T_{j-1}+1}^{t-1} Z_k} \cdot \left[\sum_{i'} (1-F+\frac{F}{b_{i't}})^{Z_t} \cdot C_{ti'} \cdot P'[X_t=i'] \right] \right\} + CSP \cdot \sum_{j=1}^{M} Z_j$$

where $C_{ti'} = C_T(i,t) + C_d(t) \cdot d_i(t)$ and CSP = cost of screening the population of suspects to the disease.

To determine an optimal policy to the screening program, there is always the possibility of enumerating all possibilities; but for a disease like breast cancer, even for a short age interval such as 40-64 and considering at most one annual screening, there are $(2)^{24} \simeq 17 \times (10)^6$ different policies to evaluate. Therefore the concept of branch and bound seems appealing. To do so, a tree is constructed which has the following characteristics:

1) There are M levels, where M is maximum number of tests per life time of an individual (at least over the time span considered).

2) At any level, for each node there are two branches corresponding to "no screening" and "screening."

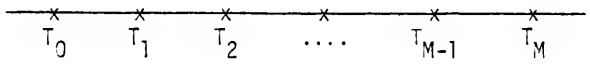
3) A node at level L is denoted by the set of M values: $\{A_1, A_2, \ldots A_M\}$, where

$$A_k = \begin{cases} 0 & \text{if screening policy is not determined} \\ 1 & \text{if } Z_k=0 \\ 2 & \text{if } Z_k=1 \end{cases}$$

Figure 26 shows the tree for the case M=3.

In order to efficiently evaluate the expected total cost of screening associated with a level L in the tree, the expression for the objective function is rederived:

Let the maximum number of screening tests per individual be M and assume that the age at which the first test could be done be denoted as T_1. The age span of an individual could be marked with T's as



where at any time T_i (i=1,...M) a test could be done, but it is not necessarily optimal to do so. Therefore, irrespective of whether or not the test is done at any of those ages, the population of suspects is divided into M groups, i.e., i-th group has an onset time somewhere between T_{i-1} and T_i. Obviously if a test is not done at age T_i, then the i-th and (i+1)st group are from the same family. Let

$$A_1 = 1 - F$$

$$B_1 = 1 - F + F/b_{it'}$$

Then for the first group, the contribution to the expected total cost would be

$$\sum_{t=T_0+1}^{T_M} W(1,t) \cdot (A_1)^{\sum_{k=T_0+1}^{t-1} Z_k} \cdot (B_1)^{Z_t} \cdot C_t$$

$$= \sum_{j=1}^{M} \sum_{t=T_{j-1}+\Delta T}^{T_j} W(1,j) \cdot (A_1)^{\sum_{k=T_0+\Delta T}^{t-\Delta T} Z_k} \cdot (B_1)^{Z_t} \cdot C_t$$

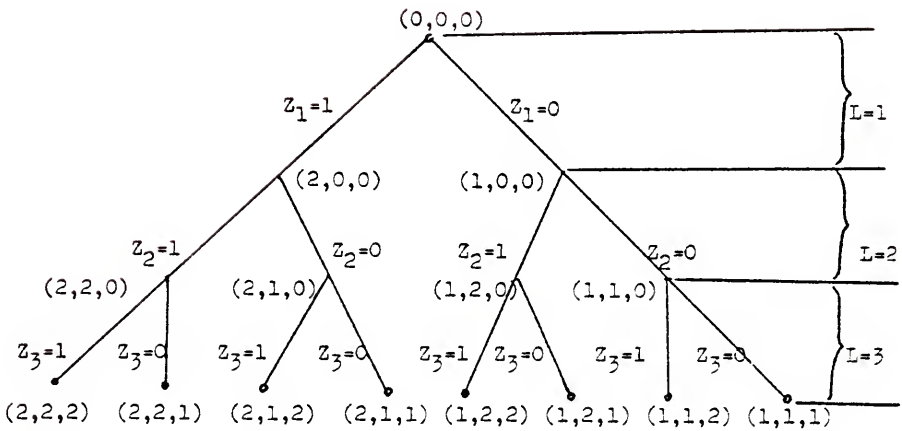


Figure 26: Construction of the tree for the case M=3.

whereas for the second group it would be:

$$\sum_{t=T_1+1}^{T_M} W(2,t) \cdot (A_1)^{\sum_{k=T_1+1}^{t-1} Z_k} \cdot (B_1)^{Z_t} \cdot C_t$$

$$= \sum_{j=2}^{M} \sum_{t=T_{j-1}+\Delta T}^{T_j} W(2,j) \cdot (A_1)^{\sum_{k=T_1+\Delta T}^{t-\Delta T} Z_k} \cdot (B_1)^{Z_t} \cdot C_t$$

The objective function consists of the contribution of all the groups:

$$OF = \sum_{i=1}^{M} \sum_{j=i}^{M} \sum_{t=T_{j-1}+\Delta T}^{T_j} \{W(i,j) \cdot (A_1)^{\sum_{k=T_{i-1}+\Delta T}^{t-\Delta T} Z_k} \cdot (B_1)^{Z_t} \cdot C_t\} +$$

$$CSP \cdot \sum_{t=1}^{M} Z_t \quad (*)$$

This cost consists of the following terms:

1) $OF_1(k)$=Expected total cost of screening for ages up to T_k.

2) $OF_2(k)$=Expected total cost of screening for ages between T_k and T_{k+1}.

3) $OF_3(k+1)$=Expected total cost of screening from T_{k+1} on.

Notice that $OF_1(k)+OF_2(k)+OF_3(k+1) = OF \; \forall \; 0 < k < M$. The expression for these three cost elements are:

$$1)\quad OF_1(k) = \sum_{i=1}^{k} \sum_{j=i}^{k} \sum_{t=T_{j-1}+\Delta T}^{T_j} \{W(i,j) \cdot (A_1)^{\sum_{\ell=T_{i-1}+\Delta T}^{t-\Delta T} Z_\ell} \cdot (B_1)^{Z_t} \cdot C_t\}$$

$$2)\quad OF_2(k) = \sum_{i=1}^{k+1} \sum_{t=T_k+\Delta T}^{T_{k+1}} \{W(i,k+1) \cdot (A_1)^{\sum_{\ell=T_{i-1}+\Delta T}^{t-\Delta T} Z_\ell} \cdot (B_1)^{Z_t} \cdot C_t\}$$

$$3)\quad OF_3(k+1) = \sum_{i=1}^{M} \sum_{j=\max(i,k+1)}^{M} \sum_{t=T_{j-1}+\Delta T}^{T_j} \{W(i,j) \cdot (A_1)^{\sum_{\ell=T_{i-1}+\Delta T}^{t-\Delta T} Z_\ell} \cdot (B_1)^{Z_t} \cdot C_t\}$$

The branch and bound approach consists of two fundamental procedures. Branching is the process of partitioning a problem into two subproblems, and bounding is the process of computing a lower bound on the optimal solution of a given subproblem.

Let P^0 denote a policy in which there are M decisions to be made. P^0 could be partitioned into two subproblems P_0^1 and P_1^1 where P_i^1 means that $Z_1=i$. Next, each of the subproblems can be partitioned into two smaller subproblems and so on.

Suppose that at some intermediate stage a complete solution has been obtained that has value S. Then any subproblem that has an associated lower bound LB > S, could be fathomed-- no more branches are generated from a fathomed branch.

The flow chart for the branch and bound procedure is shown in Figure 27. It is seen that it starts with L=1, Z_1=0; and:

1) At each level the objective function (OF) is computed and compared to S.

a) If OF < S and L < M, L=L+1, go to 1.

b) If OF $\leq$ S and L = M, store the new objective function as the new value for S, go to 2.

c) If OF > S, go to 2.

2) If Z_L=0, let Z_L=1, go to 1.

If Z_L=1, let L=L-1, Z_{L-1}=1, go to 1.

So it travels all the way through the tree - from its most right hand corner - to the bottom of the tree and then moves back and forth.

In order for a branch and bound procedure to be efficient, there must be an efficient method for computing a lower bound on the objective function at each level. Two different approaches to this problem are presented in the following section.

6.1.1 Method 1 - Discrete Varaibles

The cost-effectiveness objective function is

$$OF = \sum_{i=1}^{M} \sum_{j=i}^{M} \sum_{t=T_{j-1}+\Delta T}^{T_j} \{W(i,j) \cdot (A_1)^{\sum_{k=T_{i-1}+\Delta T}^{t-\Delta T} Z_k} \cdot (B_1)^{Z_t} \cdot C_t\} +$$

$$CSP \cdot \sum_{t=1}^{M} Z_t \quad (*)$$

In the expression for this function there are two values A_1 and B_1 where A_1=1-F and B_1=1-F+F/$b_{it'}$. Note that A_1 < 1 and B_1 > 1. In fact A_1

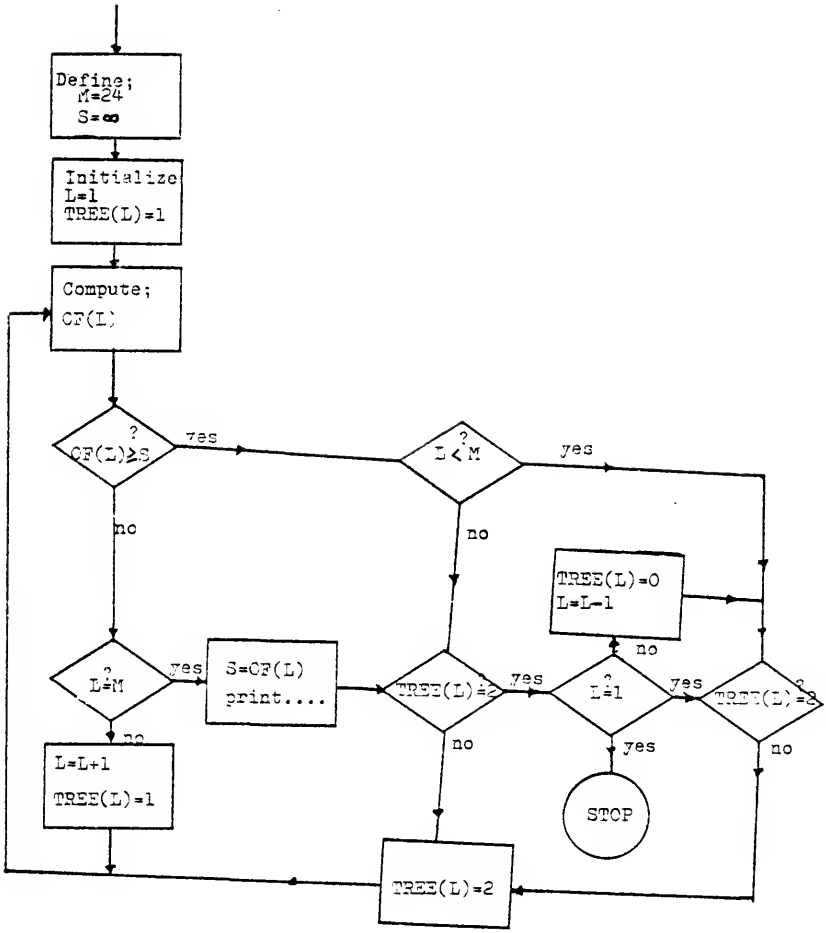


Figure 27: The flow chart for the branch and bound procedure.

is the realized reduction in the population of occult diseased individuals due to a screening test done in the past, whereas B_1 is the modified number of individuals who are detected due to a screening test at the time of screening.

The objective function could be written as

$$\sum_{i=1}^{M} \sum_{j=i}^{M} \sum_{t=T_{j-1}+DT}^{T_j} \{D_{i,j,t} \cdot A_1^{X} \cdot B_1^{Z_t}\} + CSP \cdot \sum_{j=1}^{M} Z_j$$

where $X = \sum_{k=T_{i-1}+DT}^{t-DT} Z_t$

and $D_{i,j,t} = W(i,j) \cdot C_t$.

To obtain a lower bound on this function the following quantities could be selected optimally:

$Z_t = 1 \quad \forall\ t$ (corresponding to maximum X)

$B_1 = 1$ (corresponding to minimum value for $B_1^{Z_t}$)

and $CSP = 0$ (corresponding to no charge for test itself)

$$\therefore \quad LB = \sum_{i=1}^{M} \sum_{j=i}^{M} \sum_{t=T_{j-1}+DT}^{T_j} \{W(i,j) \cdot (A_1)^{j-i} \cdot C_t\}$$

which consists of M increments associated with lower bounds on increment of objective function on i-th level, i=1,...M.

$$LB(k) = \sum_{i=1}^{k} \sum_{t=T_{k-1}+DT}^{T_k} \{W(i,k) \cdot (A_1)^{k-i} \cdot C_t\}$$

$$= \left[\sum_{i=1}^{k} W(i,k) \cdot (A_1)^{k-i}\right] \cdot \left[\sum_{t=T_{k-1}+DT}^{T_k} C_t\right]$$

$$\therefore LB = \sum_{k=1}^{M} LB(k)$$

At the beginning of the tree - no decision has yet been made - the expected total cost is assumed to be equal to the lower bound, i.e.:

$$OF(level=1) = LB = \sum_{i=1}^{k} LB(k)$$

At each step through the tree $OF_2(k)$, which is cost associated with k-th level, is computed and added to the value obtained for OF at a level before - OF(k-1). But to remove the effect of the initial lower bound, LB(k) is subtracted from the result. Mathematically,

At L=0 $\quad OF(0)=LB=LB(1)+\ldots+LB(M) = \sum_{k=1}^{M} LB(k)$

At L=1 $\quad OF(1)=LB+OF_2(1)-LB(1)$

$$=OF_2(1)+LB(2)+\ldots+LB(M)$$

$$=OF_1(1)+\sum_{k=2}^{M} LB(k)$$

AT L=2 $\quad OF(2)=OF(1)+OF_2(2)-LB(2)$

$$=OF_2(1)+OF_2(2)+LB(3)+\ldots+LB(M)$$

$$= OF_1(2)+\sum_{k=3}^{M} LB(k)$$

Therefore in general;

$$OF(k)=OF(k-1)+OF_2(k)-LB(k)$$

Starting from the first node in the tree, at each level the exact cost up to that level $OF_1(k)$ is added to a lower bound for the remaining time period $\left(\sum_{l=k}^{M} LB(l)\right)$ to obtain a lower bound on the objective function using the policy under consideration. If this value happens to be greater than a lower bound to OF already determined, the tree is fathomed.

A computer program was written that searched the tree using the above mentioned procedure. It was found that the procedure is not efficient. In fact, the value obtained for LB ($3563) was very low. Therefore it forced the search to go through most branches of the tree and increased the computational time. Table 15 shows part of computer output using this lower bound.

In a branch and bound procedure it is important to select the lower bound in such a manner that the tree can be fathomed at lower levels - a lower bound closer to the real value but less than that. To do so, it was decided to improve the lower bound. It should be noted that at any level (say L), irrespective of the (L-1) policies already made, the lower bound was taken to be a constant - this means the same lower bound is employed for policy 1 (Z_i=0, i=1,L) and policy 2 (Z_i=1, i=1,L). Obviously it does not employ the knowledge of the screening policy up to present. To get the benefit of this additional information, the following procedure was designed. Define:

LB(i,j) = Lower bound on cost going from (i-1)st to i-th level if (j-1) tests have been missed up to present time (level i-1).

For instance LB(3,2) = Lower bound on cost of going from 2nd to 3rd level if one test has been missed up to 2nd level. Let TLB(i,j) = Lower bound on cost of going from (i-1)st level on, when (j-1) tests have been

Table 15: Part of computer output using first method to compute the lower bound-Policy is: NO screen up to level L.

Level	OF_2(level)	$OF_1(L-1) + \sum_{i=L}^{M} LB(i)$	OF
1	72.95	3490.53	3563.48
2	97.17	3510.40	3607.57
3	107.26	3548.01	3670.74
4	149.43	3600.55	3749.98
5	177.13	3668.46	3845.59
6	205.66	3792.61	3958.27
7	234.86	3853.80	4088.66
8	264.58	3972.75	4237.33
9	294.65	4110.06	4404.71
10	324.89	4266.24	4591.13
11	355.16	4441.67	4796.83
12	385.30	4636.62	5021.92
13	415.14	4851.25	5266.39
14	448.51	5081.51	5530.14
15	473.26	5339.69	5812.95
16	501.23	5613.23	6114.46
17	523.25	5910.99	6434.24
18	554.16	6217.55	6771.71
19	578.80	6547.40	7126.20
20	602.00	6894.89	7496.89
21	623.61	7259.28	7882.89
22	643.42	7639.74	8283.16
23	661.40	8088.76	8750.16
24	677.25	8710.35	9387.60

missed up to (i-1)st level. It is seen that:

$$TLB(i,j) = LB(i,j) + LB(i+1,j) + \ldots + LB(M,j)$$

$$= \sum_{k=i}^{M} LB(k,j)$$

$$\therefore TLB(i,j) = TLB(i+1,j) + LB(i,j)$$

The expression for LB(i,j) would be

$$LB(i,j) = \left(\sum_{k=1}^{i} W(k,i) \cdot (A_1)^{\min \cdot (i-k,i-j)} \right) \cdot \left(\sum_{t=T_{i-1}+DT}^{T_i} C_t \right)$$

For instance: $LB(4,1)=W(1,4)A_1{}^3C_4+W(2,4)A_1{}^2C_4+W(3,4)A_1C_4+W(4,4)C_4$

whereas: $LB(4,2)=W(1,4)A_1{}^2C_4+W(2,4)A_1{}^2C_4+W(3,4)A_1C_4+W(4,4)C_4$

$LB(4,3)=W(1,4)A_1C_4+W(2,4)A_1C_4+W(3,4)A_1C_4+W(4,4)C_4$

$LB(4,4)=W(1,4)C_4+W(2,4)C_4+W(3,4)C_4+W(4,4)C_4$

The modified flow chart is very similar to the other method. Table 16 presents the new lower bounds to the objective function at each level - depending on the number of tests missed up to now. It is seen that it gets closer to the real value but the difference is still significant. Therefore the search was directed at developing a different lower bound.

Table 16: A portion of the new lower bound on objective function as a function of I and J.

TLB(i,j)	J=1	2	3	4	5	6	7	8	9	10
i = 1	3563									
2	3491	3542								
3	3437	3445	3511							
4	3378	3379	3388	3470						
5	3308	3308	3309	3321	3419					
6	3226	3226	3226	3228	3242	3357				
7	3133	3133	3133	3133	3135	3151	3285			
8	3029	3029	3029	3029	3029	3031	3050	3201		
9	2913	2913	2913	2913	2913	2913	2916	2937	3107	
10	2786	2786	2786	2786	2786	2786	2786	2789	2812	3002
.										
.										
.										

6.1.2 Method 2 - Continuous Variables

It is known in optimization theory that in an integer program problem a lower bound on the optimal solution is obtained by removing the integer constraints. Employing this concept, the variables Z_i were assumed to be continuous variables between zero and one. The objective function could be written as

$$OF = \sum_{j=1}^{M} \sum_{t=T_{j-1}+1}^{t_{max}} \left\{ W(j,t)(1-F) \prod_{k=T_{j-1}+1}^{t-1} Z_k \left[\sum_{i'} \left(1-F+\frac{F}{b_{it'}} \right)^{Z_t} C_{ti'} P'[X_t=i'] \right] \right\}$$

$$+ CSP \sum_{j=1}^{M} Z_j$$

$$= W(1,1)B_1^{Z_1} \cdot C_1 + W(1,2)A_1^{Z_1} \cdot B_1^{Z_2} \cdot C_2 + W(1,3)A_1^{Z_1+Z_2} B_1^{Z_3} C_3 + \ldots\ldots$$

$$+ W(2,2)B_1^{Z_2} \cdot C_2 \quad + W(2,3)A_1^{Z_2} \cdot B_1^{Z_3} C_3 \quad + \ldots\ldots$$

$$+ W(3,3)B_1^{Z_3} \cdot C_3 \quad + \ldots\ldots$$

$$+ CSP\ (Z_1+Z_2+\ldots\ldots+Z_M)$$

Assuming that optimization problem under consideration is unconstrained, the stationary points are found by setting the derivatives of the objective function with respect to variables Z_i to zero:

$$\frac{\partial(OF)}{\partial(Z_1)} = 0$$

$$+ \ln B_1 \cdot W(1,1) \cdot B_1^{Z_1} \cdot C_1 + \ln A_1 \cdot A_1^{Z_1} \cdot \sum_{j=2}^{n} \{W(t,j) \cdot A_1^{\sum_{k=2}^{j-1} Z_k} \cdot B_1^{Z_j} \cdot C_j\} + CSP = 0$$

$$+ Z_1 = \frac{1}{\ln\left(\frac{B_1}{A_1}\right)} \cdot \ln \left\{ - \frac{\left(\ln A_1 \cdot \sum_{j=2}^{n} W(1,j) \cdot A_1^{\sum_{k=2}^{j-1} Z_k} \cdot B_1^{Z_j} \cdot C_j\right) + \frac{CSP}{A_1^{Z_1}}}{W(1,1) \cdot C_1 \cdot \ln B_1} \right\}$$

$$\frac{\partial(OF)}{\partial(Z_2)} = 0$$

$$+ \ln B_1 \cdot W(2,2) \cdot B_1^{Z_2} \cdot C_2 + \ln A_1 \cdot A_1^{Z_2} \cdot \sum_{j=3}^{n} \{W(2,j) \cdot A_1^{\sum_{k=3}^{j-1} Z_j} \cdot B_1^{Z_j} \cdot C_j\} +$$

$$+ A_1^{Z_1} \cdot \{W(1,2)\ B_1^{Z_2} \cdot \ln B_1 \cdot C_2 + \ln A_1 \cdot A_1^{Z_2} \cdot \sum_{j=3}^{n} W(1,j) \cdot A_1^{\sum_{k=3}^{j-1} Z_k} \cdot B_1^{Z_j} \cdot C_j\} +$$

$$+ CSP = 0$$

$$+ Z_2 \frac{1}{\ln\left(\frac{B_1}{A_1}\right)} \cdot$$

$$\ln \left\{ - \frac{\ln A_1 \cdot \left[A_1^{Z_1} \cdot \sum_{j=3}^{n} W(1,j) \cdot A_1^{\sum_{k=3}^{j-1} Z_k} \cdot B_1^{Z_j} C_j + \sum_{j=3}^{n} W(2,j) \cdot A_1^{\sum_{k=3}^{j-1} Z_k} \cdot B_1^{Z_j} C_j\right] + \frac{CSP}{A_1^{Z_2}}}{[W(1,2) \cdot A_1^{Z_1} + W(2,2)] \cdot C_2 \cdot \ln B_1} \right\}$$

In general

$$Z_t = \frac{1}{\ell n\left(\frac{B_1}{A_1}\right)} \cdot \ell n$$

$$\left\{ \frac{\frac{CSP}{(A_1)^{Z_t}} + \ell n\, A_1 \cdot \left[\sum_{k=1}^{t} A_1^{\sum_{\ell=k}^{t-1} Z_\ell} \cdot \left\{ \sum_{j=t+1}^{n} W(k,j) \cdot A_1^{\sum_{\ell=t+1}^{j-1} Z_\ell} \cdot B_1^{Z_j} \cdot C_j \right\} \right]}{\left\{ \sum_{k=1}^{t} W(k,t) \cdot A_1^{\sum_{\ell=k}^{t-1} Z_\ell} \right\} \cdot C_t \cdot \ell n\, B_1} \right\}$$

Since there is a possibility some of the constraints $0 \leq Z_i \leq 1$ may be active, the Kuhn - Tucker Conditions were determined:

$$\frac{\partial(OF)}{\partial Z_t} + \lambda_t - \lambda_{t'} = 0$$

$$\lambda_t(Z_t^* - 1) = 0$$

$$\lambda_{t'}(-Z_t^*) = 0 \qquad t=1,2,\ldots.M$$

In the computer run, an iterative method was employed to search for the values of the continuous variables Z_t^* that satisfy the Kuhn - Tucker Conditions. It was found out that the convergence criterion of the iterative procedure was very slow ; this might be due to the characteristics of the cost-effectiveness objective function which is very flat in the neighborhood of the optimal solution.

Eventually it was decided to employ one of the more powerful methods of solving nonlinear programs. To do so, the "Pierre and Lowe" algorithm [75] which uses augmented Lagrangians was selected. This method is based on augmented Lagrangians and presents a comprehensive algorithm and computer code for solving constrained nonlinear programs. To make the computations more efficient, the algorithm was reduced to the case of interest - subroutines that were related to constraints were omitted - and substituted in the tree in such a manner that at each call to the evaluation of the objective function, the algorithm was called and the lower bound was computed. Table 17 shows a part of the computer output. The result was a very good lower bound to the objective function which caused the tree to fathom at such levels as 14 and 15 - almost half way through - but computation time was still large. In fact it took 30 minutes of computer time to trace the whole tree. The optimal screening policy was found to be

Z^* = screen at ages 49,51,53,55,57,58,60,61,62,...

The associated cost was $7627.59.

6.2 Search for a "Good" Heuristic Solution

Employing different methods to specify a "good" and "efficient" lower bound to the objective function failed to obtain solutions efficiently. Therefore it was decided to do a "selective enumeration process" instead. In this procedure, search for a "good" policy is done through two stages. In the first stage, the optimal or near optimal number of screening tests is determined. In stage 2, the optimal timing of the tests is determined. The procedure along with an example is described for the case of Breast Cancer.

Table 17: A part of the computer output using "Pierre and Lowe" algorithm - initially S was taken to be 7500.

L	OF (L)	TOF	Policy
1	72.	7198.	0
2	170.	7207.	0 0
3	292.	7221.	0 0 0
4	442.	7241.	0 0 0 0
5	619.	7268.	0 0 0 0 0
6	825.	7302.	0 0 0 0 0 0
7	1059.	7345.	0 0 0 0 0 0 0
8	1324.	7395.	0 0 0 0 0 0 0 0
9	1619.	7455.	0 0 0 0 0 0 0 0 0
7	1301.	7299.	0 0 0 0 0 0 1
8	1435.	7311.	0 0 0 0 0 0 1 0
9	1639.	7431.	0 0 0 0 0 0 1 0 0
10	1901.	7390.	0 0 0 0 0 0 1 0 0 0
11	2213.	7223.	0 0 0 0 0 0 1 0 0 0 0
12	2568.	7523.	0 0 0 0 0 0 1 0 0 0 0 0
12	2880.	7413.	0 0 0 0 0 0 1 0 0 0 0 1
13	3077.	7438.	0 0 0 0 0 0 1 0 0 0 0 1 0
9	1862.	7316.	0 0 0 0 0 0 1 0 1
10	2011.	7333.	0 0 0 0 0 0 1 0 1 0
. . .			

Stage I: Divide $T_{max.}$(=24) period into two equal intervals of length $(T_{max.}/2)$(=12). Within each interval there are the following $(T_{max.}/2 + 1)$ (=13) possibilities. No screen - Only one screen - ...-$T_{max}/2$ screens (=12). This produces $(T_{max.}/2 + 1)^2$ (=169) possible policies. Evaluate all those policies and determine the ε-optimal solution. For each of those policies compute the number of tests done in the first and second interval - say N(1) and N(2).

Figure 28 shows all possible policies to be evaluated at Stage I. For instance assume that N(1)=5, N(2)=8 generated the best solution within all 169 possible policies.

Stage II-a: Divide each interval into two subintervals. This generates a total of four subintervals of length $T_{max.}/4$ (=6). For the first two subintervals, generate all feasible subpolicies in which there are I_1 tests in the first subinterval and J_1 test(s) in the second one - such that I_1+J_1=N(1). For any such subpolicy, generate all feasible policies in which there are I_2 test(s) in the third subinterval and J_2 test(s) in the fourth one - such that I_2+J_2=N(2). Determine all ε-optimal policies. Let the number of tests in any subinterval in the optimal solution be L(i), i=1,...4.

Figure 29 shows the feasible policies for the example problem. It is seen that there are a total of 6 x 5 = 30 possibilities. Assume that L(1)=4, L(2)=1, L(3)=6, L(4)=8.

Stage II-b: Divide each interval into two equal subintervals. This generates a total of eight subintervals of length $T_{max.}/8$ (=3). Do the same procedure done in Stage II-a for L(1), L(2), L(3) and L(4). Determine ε-optimal solutions and let the optimal number of test(s) in each subinterval be K(i),i=1,...8.

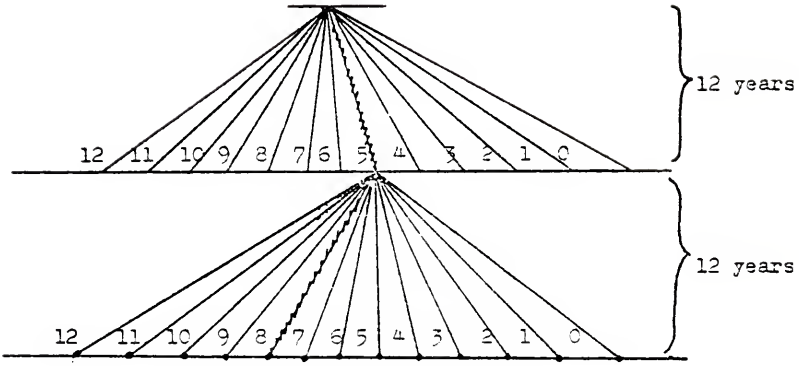


Figure 28: All possible policies for stage I.

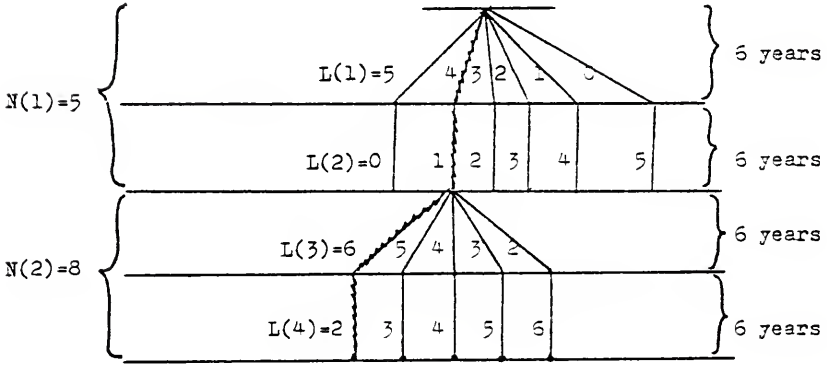


Figure 29: Stage II-a for the example problem.

Figure 30 shows the feasible policies for the example problem. It is seen that there are a total of 3.2.1.3=18 possibilities. Assume that K(i)=2,2,1,0,3,3,1 and 1.

Stage II-c: Divide each interval into three subintervals of unit length. For the i-th interval generate all subpolicies that have a total of K(i) test(s). The optimal solution to this stage could be used as a "good" solution to the optimization problem.

In the case of the example problem all possible solutions are shown in Figure 31. Note that for K(i)=0 or 3 there is only one policy, whereas for K(i)=1 or 2 there are three different policies.

Since the solution to this heuristic method depends on the location of the tests within an interval in Stages I, II-a and II-b, in the computer runs that were made, these locations were varied in a random fashion and different solutions were obtained. The best solution was a very perculiar one:

$$Z_t = 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1$$

which means; do screening every three years up to age 50,
then do screening every other year up to age 53,
and do screening annually from then on.

The pattern of screening is in agreement with the optimal solutions obtained by other investigators.

Table 18 which compares this solution with some of the conventional solutions and the optimal solution shows the ability of the method in determination of a "good" policy - 1% reduction was obtained over the best conventional solution. Comparison with optimal solution shows that the heuristic method generated a solution which was within 0.07%(*) of

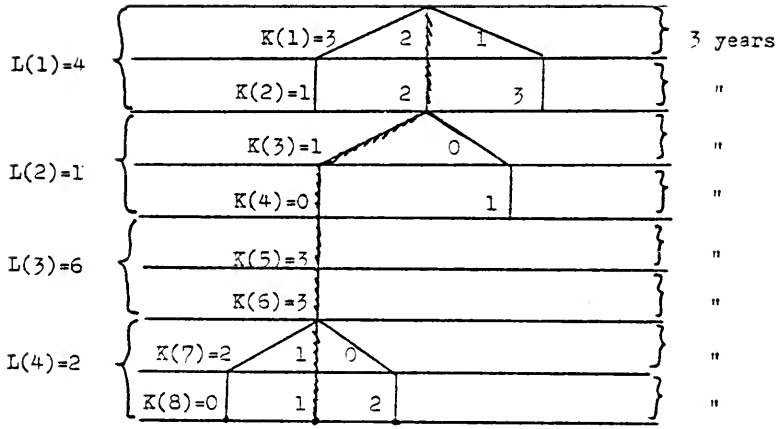


Figure 30: Stage II-b for the example problem.

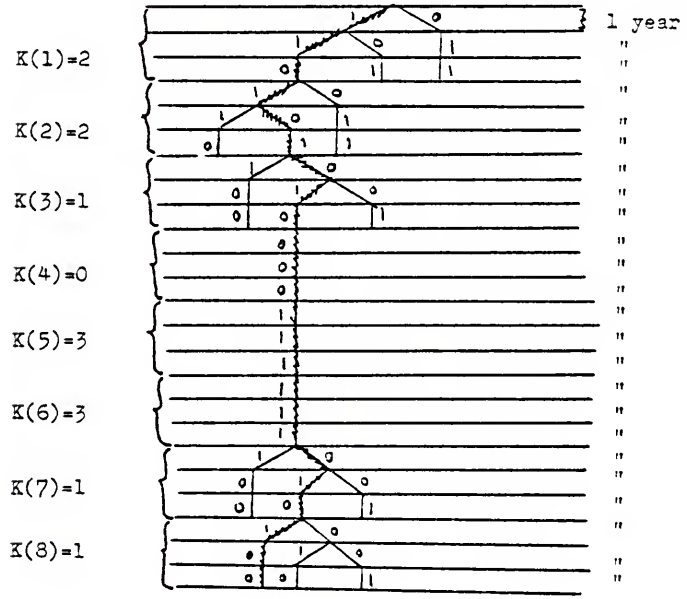


Figure 31: Stage II-c for the example problem.

Table 18: Comparison between heuristic, optimal solution and conventional solutions.

Policy	Age of Screen	Associated Cost	Cost as a Percent of Optimum
Optimal Policy	49,51,53,55,57,58,60,61,62	7627.59	100.
Heuristic Method	43,46,49,51,53 on	7633.13	100.07
No screen	-	9387.63	123.07
One screen	53	8993.10	117.90
Two screens	46,58	8731.52	114.47
Three screens	46,53,60	8403.18	110.16
Four screens	45,50,55,60	8182.41	107.27
Five screens	45,49,53,57,61	7978.27	104.59
Six screens	41.45,49,53,57,61	7993.38	104.79
Twelve screens	Every two years	7698.17	100.92
24 screens	Every year	8013.58	105.06

the optimal solution - it took 29 seconds of computer time to generate such a solution. Table 18 also shows that there are many different policies with the same associated cost. This property of the cost-effectiveness objective functions makes the heuristic solution a very good approximation to the optimal policy.

The same method can also be used in the case of screening of diseases when the total number of screening tests during each age period is taken to be a constant and only the interval between screening examinations is important.

(*) $\varepsilon=(7633.13-7627.59)/7627.59=0.07$

CHAPTER SEVEN

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This research presented a mathematical model of the cancer diseases based on the hypothesis that the prognosis is a function of stage of the disease and age of the individual at screening time. The model employs the data on the natural probability of being in a detected state and modifies it to the case of any screening policy. It develops insights into the process of screening of progressive diseases. The model described in this research provides the decision makers with a means for deciding on the number of tests given to an individual and the associated ages at screenings.

The model was implemented for the case of Neuroblastoma and breast cancer. For these diseases, the benefits of different screening policies were evaluated under different assumptions about the reliability of the screening techniques and the form of disease progression in an individual. The model was also validated against reported data and the results were compared with other models.

The general conclusions were

1) Within a reasonable range of test cost, it is normally beneficial to do screening examinations.

2) The optimal screening strategy is not usually periodic and the screening interval is a function of the age of individual, but periodic screenings are close to optimal.

3) In determination of an optimal policy it should be noted that there are different policies which have the same measures of effectiveness.

4) In measuring the performance of a new screening program, the transient effect should be considered because the prevalence pool decreases to its steady state value as a function of time.

The model presented is distinct from others in the following respects.

1) It is built on the basis of such parameters as natural probability of detection, whose estimation is easier than most of the other commonly used parameters.

2) It offers values of such parameters of interest as the probability of death due to the disease, probability of being detected at a particular examination and probability of being surfaced between two successive tests.

3) Finally a variety of different problems related to the screening process such as dependency between examination results and time dependency of the screening program can be represented.

The author would like to suggest some potentially useful topics for future research:

1) In this research irrespective of the progression of the disease in the detected states it was assumed that the individual either dies or survives. An extension of this research would be investigation of an appropriate model for the case of recurrent diseases in which treatment of the individual may only postpone the time of death.

2) The cost-effectiveness objective function was shown to be a nonlinear integer problem. An investigation could be made to determine

the closed form expression for optimal solution to the linear approximation to the objective function.

3) In this research it was assumed that the test itself is harmless, whereas in case of breast cancer Knox [55,56] has shown that screening may cause cancer. A generalization of the presented model could be made to include those possibilities.

4) Another assumption employed in this research was that the test is so designed that there is no chance of mistake in stage recognition. It would be interesting to relax this assumption and determine the expression under the more general case.

5) In Section 5.5 of this research the model was generalized to take into account the fact that some of the individuals may not participate in all the screens. An extension of this generalization would be to consider a population which is not closed, i.e., some individuals come from another population with different screening programs and some individuals leave the population forever.

6) In this research there was no provision for a treated individual to enter the population again. The model could be generalized to take different risk level individuals with different onset distributions, probability of natural detection, interest in being screened, self-examination and the reliability of the test.

7) Finally it seems reasonable to modify the model to take data from a previous screening program and modify it to the case of a new program.

APPENDIX

DERIVATION OF THE GENERAL EXPRESSION FOR TRANSIENT PROBLEM

$$\text{Let } A_t = \sum_{i'} \left(1-F+\frac{F}{b_{it}}\right)^{Z_t}\cdot[C_T(i,t)+C_d(t)\cdot d_i(t)]\ P'[X_t=i']$$

and let (i,j) denote groups i in the j-th interval, where

The ith group is composed of those individuals who had an onset sometime between the (i-1)st and the i-th screen.

The jth interval is a time interval between $\hat{T}_j-\hat{T}_1$ and $\hat{T}_{j+1}-\hat{T}_i$.

In the first interval j=1, the expression would be

$$\begin{array}{llll}
i=1 & \sum_{t=1}^{\hat{T}_1} A_t + \sum_{\theta=1}^{M-1}\left(\sum_{t=\hat{T}_\theta+1}^{\hat{T}_\theta+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_\theta+T+1}^{\hat{T}_\theta+1} A_t\right) & + \sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty} A_t & \\
j=1, & & & \\
i=2 & \sum_{t=\hat{T}_1+1}^{\hat{T}_2} A_t+\sum_{\theta=2}^{M-1}\left(\sum_{t=\hat{T}_\theta+1}^{\hat{T}_\theta+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_\theta+T+1}^{\hat{T}_\theta+1} A_t\right) & + \sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty} A_t & \\
j=1, & & & \\
j=3 & \sum_{t=\hat{T}_2+1}^{\hat{T}_3} A_t+\sum_{\theta=3}^{M-1}\left(\sum_{t=\hat{T}_\theta+1}^{\hat{T}_\theta+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_\theta+T+1}^{\hat{T}_\theta+1} A_t\right) & + \sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty} A_t & \\
j=1, & & & \\
i=4 & \sum_{t=\hat{T}_3+1}^{\hat{T}_4} A_t + \quad 0 & + \sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty} A_t & \\
j=1, & & & \\
i=5 & 0 \quad + \quad 0 & + \quad 0 \quad + \sum_{t=\hat{T}_M+1}^{t_\infty} A_t & \\
j=1, & & &
\end{array}$$

In the second interval, j=2, (i.e., $\hat{T}_2-\hat{T}_1 < \hat{T} \leq \hat{T}_3-\hat{T}_1$)

$$i=1,\ j=2,\quad \sum_{\theta=1}^{2}\left(\sum_{t=\hat{T}_{\theta-1}+1}^{\hat{T}_\theta}(1-F)^{\theta-1}\cdot A_t\right)+\sum_{\theta=2}^{M-1}\left(\sum_{t=\hat{T}_\theta+1}^{\hat{T}_{\theta-2}+T}(1-F)^2\cdot A_t+\sum_{t=\hat{T}_{\theta-1}+T+1}^{\hat{T}_{\theta+1}}(1-F)\cdot A_t\right)+$$

$$+\left(\sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)^2\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{\hat{T}_M+T}(1-F)A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty}A_t\right)$$

$$i=2,\ j=2,\quad \sum_{\theta=1}^{2}\left(\sum_{t=\hat{T}_\theta+1}^{\hat{T}_{\theta+1}}(1-F)^{\theta-1}\cdot A_t\right)+\sum_{\theta=3}^{M-1}\left(\sum_{t=\hat{T}_\theta+1}^{\hat{T}_{\theta-2}+T}(1-F)^2\cdot A_t+\sum_{t=\hat{T}_{\theta-1}+T+1}^{\hat{T}_{\theta+1}}(1-F)\cdot A_t\right)+$$

$$+\left(\sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)^2\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty}A_t\right)$$

$$i=3,\ j=2,\quad \sum_{\theta=1}^{2}\left(\sum_{t=\hat{T}_{\theta+1}+1}^{\hat{T}_{\theta+2}}(1-F)^{\theta-1}\cdot A_t\right)+0+\left(\sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)^2\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty}A_t\right)$$

$$i=4,\ j=2,\quad \sum_{\theta=1}^{2}\left(\sum_{t=\hat{T}_{\theta+2}+1}^{\hat{T}_{\theta+3}\leq T_M}(1-F)^{\theta-1}\cdot A_t\right)+0+\sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty}A_t$$

$$i=5,\ j=2,\quad 0+0+0+\sum_{t=\hat{T}_M+1}^{t_\infty}A_t$$

In the third interval, j=3, (i.e., $\hat{T}_3-\hat{T}_1 < T \leq \hat{T}_4-\hat{T}_1$)

$$i=1,\ j=3,\quad \sum_{\theta=1}^{3}\left(\sum_{t=\hat{T}_{\theta-1}+1}^{\hat{T}_\theta}(1-F)^{\theta-1}\cdot A_t\right)+\sum_{\theta=3}^{M-1}\left(\sum_{t=\hat{T}_\theta+1}^{\hat{T}_{\theta-2}+T}(1-F)^3\cdot A_t+\sum_{t=\hat{T}_{\theta-2}+T+1}^{\hat{T}_{\theta+1}}(1-F)^2\cdot A_t\right)+$$

$$\left(\sum_{t=\hat{T}_M+1}^{\hat{T}_{M-2}+T}(1-F)^3\cdot A_t+\sum_{t=\hat{T}_{M-2}+T+1}^{\hat{T}_{M-1}+T}(1-F)^2\cdot A_t+\sum_{t=\hat{T}_{M-1}+T+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty}A_t\right)$$

$$i=2,\ j=3,\quad \sum_{\theta=1}^{3}\left(\sum_{t=\hat{T}_{\theta}+1}^{\hat{T}_{\theta+1}}(1-F)^{\theta-1}\cdot A_t\right)+0+\left(\sum_{t=\hat{T}_M+1}^{\hat{T}_{M-2}+T}(1-F)^{3}\cdot A_t+\sum_{t=\hat{T}_{M-2}+T+1}^{\hat{T}_{M-1}+T}(1-F)^{2}\cdot A_t+\sum_{t=\hat{T}_{M-1}+T+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty}A_t\right)$$

$$i=3,\ j=3,\quad \sum_{\theta=1}^{3}\left(\sum_{t=\hat{T}_{\theta+1}+1}^{\hat{T}_{\theta+2}\le T_M}(1-F)^{\theta-1}\cdot A_t\right)+0+0+\sum_{t=\hat{T}_M+1}^{\hat{T}_{M-1}+T}(1-F)^{2}\cdot A_t+\sum_{t=\hat{T}_{M-1}+T+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty}A_t$$

$$i=4,\ j=3,\quad \sum_{\theta=1}^{3}\left(\sum_{t=\hat{T}_{\theta+2}+1}^{\hat{T}_{\theta+3}\le T_M}(1-F)^{\theta-1}\cdot A_t\right)+0+0+\sum_{t=\hat{T}_M+1}^{\hat{T}_M+T}(1-F)\cdot A_t+\sum_{t=\hat{T}_M+T+1}^{t_\infty}A_t$$

$$i=5,\ j=3,\quad 0+0+0+0+\sum_{t=\hat{T}_M+1}^{t_\infty}A_t$$

A similar analysis gives the expressions for j=4 and j=5. Therefore for i-th group in the j-th interval the expected total treatment and death cost in the j-th interval is

$$O.F.(j)=\sum_{i=1}^{M+1}W(i,j)]\cdot[O.F.(i,j)] \qquad j\le M+1$$

$$O.F.(j)=\sum_{i=1}^{M+1}\left\{\sum_{\theta=1}^{j}\left(\sum_{t=\hat{T}_{(\theta+i-2)}+1}^{\text{up to }\hat{T}_{(\theta+i-1)}\le\hat{T}_M}W(i,t)\cdot(1-F)^{\theta-1}\cdot A_t\right)+\sum_{\theta=i+j}^{M}\left(\sum_{t=\hat{T}_{(\theta-1)}+1}^{\hat{T}_{(\theta-j)}+T}W(i,t)(1-F)^{j}\cdot A_t+\sum_{t=\hat{T}_{(\theta-j)}+T+1}^{\hat{T}_\theta}W(i,t)(1-F)^{j-1}\cdot A_t\right)+\right.$$

$$
\left\{
\begin{array}{l}
\displaystyle\sum_{\theta=1}^{j+1}\left(\sum_{t=\hat{T}_M+1,\,\hat{T}_{(M-j+\theta-1)}+T+1}^{\min\{\hat{T}_{M-j+\theta}+T,\,t_\infty\}} \hat{W}(i,t)(1-F)^{j-\theta+1}\cdot A_t\right) \\
\text{or} \\
\displaystyle\sum_{\theta=1}^{M-i+2}\left(\sum_{t=\hat{T}_M+1,\,\hat{T}_{(i+\theta-2)}+T+1}^{\min\{T_{(i+\theta-1)}+T,\,t_\infty\}} W(i,t)\cdot(1-F)^{M-i+2-\theta}\cdot A_t\right)
\end{array}
\right.
\begin{array}{l}
\text{if } i+j\leq M+1 \\
\text{independent} \\
\text{of } i \\
\\
\text{if } i+j\geq M+1 \\
\text{independent} \\
\text{of } j
\end{array}
$$

"Notation" : In this formulation $\sum_{t=a,b}^{\min(c,d)} L(t)$ means: the term starts with a and the next terms from b. Moreover the upper limit of the summation is the minimum value of c and d, i.e., stop whenever the upper bound is reached.

LIST OF REFERENCES

[1] L. T. Adams, "Neuroblastoma, review of literature and report of eighteen cases," North Carolina Medical Journal 27, 113-125 (1966).

[2] A. Albert, P. M. Gertman, and T. A. Louis, "Screening for the early detection of cancer. I - Temporal natural history of a progressive disease state," Research Report 20-75, Boston Uni., 1976.

[3] A. Albert, P. M. Gertman, T. A. Louis, and S. Lou, "Screening for the early detection of cancer. II - The impact of screening on the natural history of the disease," Research Report 7-76, Boston Uni., 1977.

[4] P. Armitage, C. M. McPherson, and J. C. Copes, "Statistical studies of prognosis in advanced breast cancer," J. of Chronic Dis. 22, 343-360 (1969).

[5] L. M. Axtell, S. J. Cutler, and M. H. Myers, End results in cancer report No. 4, DHEW publication No. 73-272 (NIH), 1972.

[6] B. A. Barron and R. M. Richart, "A statistical model of the natural history of cervical carcinoma based on a prospective study of 557 cases," J. Nat. Cancer Inst. 41, 1343-1353 (1968).

[7] J. B. Beckwith and R. F. Martin, "Observations on the histopathology of neuroblastoma," J. of Ped. Surg. 3, 106-110 (1968).

[8] J. B. Beckwith and E. Perrin, "In situ neuroblastoma, A contribution to the natural history of neural crest tumors," Amer. J. of Path. 43, 1089-1104 (1963).

[9] J. W. Berg and G. Robbins, "The failure of a model to predict cancer survival," J. of Chronic Dis. 20, 809-814 (1967).

[10] L. Blumenson, "When is screening effective in reducing the death rate? " Math. Biostat. 30, 273-303 (1976).

[11] L. Blumenson and I. Bross, "A mathematical analysis of the growth and spread of breast cancer," Biometrics 25, 95-109 (1969).

[12] M. Bodian, "Neuroblastoma," Ped. Clin. N.A. 6, 449-472 (1959).

[13] N. Breslow and B. McCann, "Statistical estimation of prognosis for children with neuroblastoma," Cancer Res. 31, 2098-2103 (1971).

[14] I. Bross and L. Blumenson, "Statistical testing of a deep mathematical model for breast cancer," J. of Chronic Dis. 21, 493-506 (1968).

[15] I. Bross, L. Blumenson, N. H. Slack, and R. L. Priore, "A two disease model for breast cancer," in Prognostic factors in breast cancer, A. P. M. Forrest and P. B. Kumkler (Eds.) Baltimore, Williams and Wilkims Co., 288-300 (1968).

[16] M. M. Copeland, "American joint committee on cancer staging and end results reporting," Cancer 18, 1637-1640 (1965).

[17] S. J. Cutler and M. Myers, "Clinical classification of extent of disease in cancer of the breast," J. Nat. Cancer Inst. 39, 193-203 (1967).

[18] H. W. Daregon, "Neuroblastoma," J. of Ped. 61, 456-471 (1962).

[19] E. Day and F. Venet, "Periodic cancer detection examination as a cancer control measure," in Fourth national cancer conference proceedings 61, 705-709 (1961).

[20] A. A. De Lormer, K. U. Bragg, and G. Linden, "Neuroblastoma in childhood," Amer. J. of Dis. Child. 118, 441-450 (1969).

[21] W. Dinicola and N. Movassaghi, "Prognosis in black children with neuroblastoma," Cancer 36, 1151-1153 (1975).

[22] R. Eker, "Some prognostic factors for carcinomas of the colon and rectum," Acta Chir. Scan. 126, 636-656 (1963).

[23] A. E. Evans, "Staging and treatment of neuroblastoma," in Clinical pediatrics oncology, M. M. Vuksanovic (Ed.) New York, Futura Co., 89-91 (1972).

[24] A. E. Evans, V. Albo, G. J. D'Angio, J. Z. Finklestein, S. Leiken, T. Santulli, J. Weiner, and C. D. Hammond, "Factors influencing survival of children with non metastatic neuroblastoma," Cancer 38, 661-666 (1971).

[25] A. E. Evans, J. Blore, R. Hadley, and S. Tanindi, "The Labrosse spot test: A practical aid in the diagnosis and management of children with neuroblastoma," Pediat. 47, 913-915 (1971).

[26] A. E. Evans, G. J. D'Angio, and C. E. Koop, "Diagnosis and treatment of neuroblastoma," Ped. Clin. N.A. 23, 161-170 (1976).

[27] A. E. Evans, G. J. D'Angio, and J. Randolph, "A proposed staging for children with neuroblastoma," Cancer 27, 374-378 (1971).

[28] T. C. Everson and W. H. Cole, Spontaneous regression of cancers Philadelphia, Saunders Co., 1966.

[29] M. Feinleib, "The stable disease model (Abstract)," Biometrics 23, 1299 (1967).

[30] R. M. Filler, D. G. Traggis, N. Jaffe and G. F. Vawter, "Favorable outlook for children with mediastinal neuroblastoma," J. of Ped. Surg. 7, 136-143 (1972).

[31] R. Fink, S. Shapiro, and J. Lewison, "The reluctant participant in a breast cancer screening program," Public Health Rep. 83, 479-490 (1968).

[32] R. Fink, S. Shapiro, and R. Roester, "Impact of efforts to increase participation in repetitive screening for early breast cancer," Amer. J. of Public Health 62, 328-336 (1972).

[33] J. Fortner, A. Nicastri, and M. L. Murphy, "Neuroblastoma--Natural history and results of treating 133 cases," Ann. Surg. 167, 132-142 (1968).

[34] I. G. Furnival, "Accuracy of screening methods for the diagnosis of breast cancer," British Med. J. 4, 461-463 (1970).

[35] H. P. Galliher, "Cost-effective planned lifetime schedules of pap smears," Michigan Cancer Foundation, March 1976.

[36] V. Gilbersten, "Improved prognosis for survival following earlier detection of cancer," Geriatrics 25, 149-154 (1970).

[37] S. E. Gitlow, L. M. Bertani, S. W. Dziedzic, B. L. Wong, and L. Rozycki, "Relationship of catecholamine catabolite excretion to prognosis in neuroblastoma," Amer. Ped. Program and Abst. 216 (1971).

[38] S. E. Gitlow, L. M. Bertani, A. Rausen, D. Gribetz, and S. W. Dziedzic, "Diagnosis of Neuroblastoma by qualitative and quantitative determination of catecholamine metabolites in urine," Cancer 25, 1377-1383 (1970).

[39] J. D. Goldberg and J. T. Wittes, "The estimation of false negatives in medical screening," in process of being published in Biometrics 1976.

[40] R. E. Greenberg and L. I. Gardner, "New diagnostic test for neural tumors of infancy," Pediat. 24, 683-684 (1959).

[41] M. Greer, A. H. Anton, C. M. Williams, and C. A. Echevarria, "Tumors of neural crest origin," Arch. Neurol. (Chicago) 13, 139-148 (1965).

[42] R. E. Gross, S. Farber, and L. W. Martin, "Neuroblastoma sympatheticum a study and report of 217 cases," Pediat. 23, 1179-1191 (1952).

[43] G. H. Guin, E. F. Gilbert, and B. Jones, "Incidental neuroblastoma in infants," Amer. J. of Clin. Path. 51, 126-136 (1969).

[44] G. W. Hallett, "Urine VMA screening for neuroblastoma--Is it worth the cost? " Pediat. 51, 757 (1973).

[45] J. B. Hastings, "Mass screening for colorectal cancer," Amer. J. Surg. 127, 228-233 (1974).

[46] L. Helson, "Neuroblastoma, early diagnosis a key to successful treatment," Ped. Annals 3, 46-54 (1974).

[47] P. Hinton and F. Buschke, "Neuroblastoma in children, 42 cases," Radiol. Clin. Biol. 37, 19-28 (1968).

[48] R. C. Horn, C. E. Koop, and W. B. Kiesewetter, "Neuroblastoma in children," Lab. Invest. 5, 106-119 (1956).

[49] G. Hutchison and S. Shapiro, "Lead time gained by diagnostic screening for breast cancer," J. Nat. Cancer Inst. 41, 665-681 (1968).

[50] D. H. James, Jr., "Proposed classification of neuroblastoma," J. of Ped. 71, 764 (1967).

[51] L. M. Kinnier Wilson and C. J. Draper, "Neuroblastoma, its natural history and prognosis," British Med. J. 3, 301-307 (1974).

[52] R. Kirch and M. Klein, "Examination schedules for breast cancer," Cancer 33, 1444-1450 (1974).

[53] M. Klein and R. Kirch, "Surveillance schedules for medical examination," Management Sci. 20, 1403-1409 (1974).

[54] H. E. Knoerschild, "Sigmoidoscopic examination; Routine test for patients in hospital," California Med. 102, 359-360 (1965).

[55] E. G. Knox, "A simulation system for screening procedures," in The future and present indicatives, ninth series, G. McLachlan (Ed.) Oxford, University Press, 19-55 (1973).

[56] E. G. Knox, "Simulation studies of breast cancer screening programmes," Health Services Research Center, Uni. of Birmingham, 1974.

[57] D. Kodlin, "A note on the cost-benefits problem in screening for breast cancer," Methods of Information in Med. 11, 242-247 (1972).

[58] C. E. Koop, "Nephroblastoma and Neuroblastoma in children," Amer. J. of Surg. 101, 566-570 (1961).

[59] C. E. Koop, "Current management of nephroblastoma and neuroblastoma," Amer. J. of Surg. 107, 497-501 (1964).

[60] C. E. Koop and J. R. Hernandez, "Neuroblastoma; Experience with 100 cases in children," Surgery 56, 726-733 (1964).

[61] C. E. Koop and D. G. Johnson, "Neuroblastoma; An assessment of therapy in reference to staging," J. of Ped. Surg. 6, 595-600 (1971).

[62] E. H. Labrosse, "Biochemical diagnosis of neuroblastoma; Use of a urine spot test," Proc. of the Amer. Ass. for Cancer Res. 9, 39 (1968).

[63] J. A. H. Lee, "Summer and death from neuroblastoma," British Med. J. 2, 404-407 (1967).

[64] A. S. Leonard, S. A. Roback, M. E. Nesbit, and E. Freier, "The VMA test strip; A new tool for mass screening," J. of Ped. Surg. 7, 528-531 (1972).

[65] T. Lincoln and G. Weiss, "A statistical evaluation of recurrent medical examinations," Oper. Res. 12, 187-205 (1964).

[66] T. A. Louis, A. Albert, and S. Heghinian, "Screening for the early detection of cancer. III - Estimation of disease natural history," Boston Uni., 1976.

[67] J. Makinen, "Microscopic patterns as a guide to prognosis of neuroblastoma in childhood," Cancer 29, 1637-1646 (1972).

[68] N. Mantel, "Evaluation of a class of diagnostic tests," Biometrics 7, 240-246 (1951).

[69] M. H. Mayers, L. M. Axtell, and M. Zelen, "The use of prognostic factors in predicting survival for breast cancer patients," J. of Chronic Dis. 19, 923-933 (1966).

[70] R. M. McAllister, "Neuroblastoma; A viral etiology? " J. of Ped. Surg. 3, 138-141 (1968).

[71] D. E. Miller, "What is early diagnosis doing? " Cancer 37, 426-432 (1976).

[72] R. W. Miller, "Fifty-two forms of childhood cancer," J. of Ped. 75, 685-689 (1969).

[73] R. W. Miller, J. F. Fraumeni, and J. A. Hill, "Neuroblastoma; epidemiologic approach to its origin," Amer. J. of Dis. Child. 115, 253-261 (1968).

[74] D. R. Peterson, A. H. Bill, and I. S. Kirkland, "Neuroblastoma trends in time," J. of Ped. Surg. 4, 244-249 (1969).

[75] D. A. Pierre and M. J. Lowe, Mathematical programming via augmented Lagrangians, Reading Massachusetts, Addison-Wesley Co., 1975.

[76] D. Pinkel, C. Pratt, C. Holton, D. James, E. Wrenn, and O. Hustu, "Survival of children with neuroblastoma treated with combination chemotherapy," J. of Ped. 73, 928-931 (1968).

[77] C. J. Priebe and H. W. Clatworth, "Evaluation of the treatment of 90 children," Archives of Surgery 95, 538-544 (1967).

[78] P. Prorok, "The theory of periodic screening I; Lead time and proportion detected," Advanced App. Prob. 8, 127-143 (1976).

[79] P. Prorok, "The theory of periodic screening II; Doubly bounded recurrence times and mean lead time and detection probabilities estimation," Advanced App. Prob. 8, 460-476 (1976).

[80] D. Rice, Estimating the cost of illness, U.S. DHEW Public Health Service, Health Economics Series No. 6, 1976.

[81] A. Samitsky and F. Desposito, "A survey of American experience with vitamin B-12 therapy of neuroblastoma," J. of Ped. 67, 99-103 (1965).

[82] W. Sawitz and B. D. Karpinos, "Statistical problems involved in application of NIH swab for the diagnosis of oxyuriasis," Amer. J. Hyg. 35, 15-26 (1942).

[83] D. B. Shahon, B. T. Santoro, and O. H. Wangensteen, "Periodic examinations of the breast cancer," J. Nat. Cancer Conference Proceedings 4, 225-228 (1961).

[84] S. Shapiro, J. Goldberg, and G. Hutchinson, "Lead time in breast cancer detection and implications for periodicity of screening," Amer. J. of epid. 100, 357-366 (1974).

[85] S. Shapiro, P. Strax, and L. Venet, "Periodic breast cancer screening" in the pre-symptomatic detection and early diagnosis, C. L. E. H. Sharp and H. Keen (Eds.) Baltimore, Williams and Wilkins Co., 203-236 (1968).

[86] S. Shapiro, P. Strax, and L. Venet, "Periodic breast cancer screening in reducing mortality from breast cancer," J.A.M.A. 215, 1777-1785 (1971).

[87] S. Shapiro, P. Strax, L. Venet, and W. Venet, "Changes in five year breast cancer screening program," Seventh Nat. Cancer Conference Proceedings 73, 663-678 (1973).

[88] M. Shwartz, "Analysis of screening strategies for breast cancer; A review of literature," Michigan Cancer Foundation, March 1976.

[89] M. Shwartz, "Analysis of the benefits of serial screening to detect early breast cancer," Michigan Cancer Foundation, March 1976.

[90] M. Shwartz, "Analysis of the benefits of serial screening to detect early breast cancer," Michigan Cancer Foundation, August 1976.

[91] M. Shwartz, "An analysis of the benefits of serial screening for breast cancer based on a mathematical model of the disease, Part II-1," Michigan Cancer Foundation, May 1977.

[92] M. Shwartz, "A mathematical model used to analyze breast cancer screening strategies, Part II-2," Michigan Cancer Foundation, May 1977.

[93] M. Shwartz, "Analysis of screening strategies for breast cancer. Supplementary material, Part II-3," Michigan Cancer Foundation, May, 1977.

[94] J. G. Stella, O. Schweisguth, and M. Schlienger, "Neuroblastoma," Amer. J. of Roen. Radium Therapy and Nuclear Med. 108, 324-332 (1970).

[95] P. Strax, L. Venet, S. Shapiro, and S. Gross, "Mammography and clinical examination in mass screening for cancer of the breast," Cancer 20, 2184-2188 (1967).

[96] W. W. Sutow, E. A. Gehan, R. M. Heyn, F. H. Kung, R. W. Miller, M. L. Murphy, and D. G. Traggis, "Comparison of survival curves 1956 versus 1962," Pediat. 48, 800-811 (1970).

[97] G. M. Tallis and G. Sarfaty, "On the distribution of the time to reporting cancers with applications to breast cancer in women," Math. Biosciences 19, 371-376 (1974).

[98] D. Thompson and R. Disney, "A mathematical model of progressive disease and screening," VRI, November 1976.

[99] D. Thompson and T. C. Doyle, "Analysis of screening strategies for colorectal cancer," Michigan Cancer Foundation, April 1976.

[100] G. F. Vawter, "Prognosis of neuroblastoma," in Clinical pediatrics and oncology, M. M. Vuksanovic (Ed.) New York, Future Co., 93-97 (1972).

[101] L. Venet, P. Strax, W. Venet, and S. Shapiro, "Adequencies and inadequencies of breast examinations by physicians in mass screenings," Cancer 28, 1546-1551 (1971).

[102] M. L. Voorhess, "Neuroblastoma with normal urinary catecholamine excretion," J. of Ped. 78, 680-683 (1971).

[103] M. L. Voorhess and L. I. Gardner, "Studies of catecholamine excretion by children with neural tumors," J. Clin. Endoc. Meta. 22, 126-133 (1962).

[104] M. L. Voorhess and L. I. Gardner, "The value of serial determination in children with neuroblastoma," Pediat. 30, 241-246 (1962).

[105] P. A. Voute, W. J. Van Pulton, and J. M. V. Burgers, "Tumors of sympathetic nervous systems," in Cancer in children, H. J. G. Bloom (Ed.) New York, Springer-Verlag Co., 138-148 (1975).

[106] G. H. Weiss and T. L. Lincoln, "Analysis of repeated examinations for detection of occult disease," Health Ser. Res. 1, 272-286 (1966).

[107] J. Whang-Peng and J. M. Bennett, "Cytogenic studies in metastatic neuroblastoma," Amer. J. of Dis. Child. 115, 703-708 (1968).

[108] T. E. Williams and M. H. Donaldson, "Neuroblastoma" in Clinical pediatric oncology, W. W. Sutow (Ed.) New York, Mosby Co., 384-310 (1972).

[109] J. T. Wittes, "Notes on the bias and estimated variance of Chapman's two sample capture-recapture population estimate," Biometrics 28, 592-597 (1972).

[110] J. T. Wittes, "Applications of a multinomial capture-recapture model to epidemiological data," J. Amer. Stat. Ass. 69, 93-97 (1974).

[111] J. T. Wittes, T. Colton, and V. Sidel, "Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources," J. of Chronic Dis. 27, 25-36 (1974).

[112] J. T. Wittes and V. Sidel, "A generalization of the sample capture-recapture model with applications to epidemiological researches," J. of Chronic Dis. 21, 287-301 (1968).

[113] M. Zelen, "Problems in the early detection of disease and the finding of faults," Bulletin of the Inter. Stat. Ins. Proceedings 44, 649-661 (1972).

[114] M. Zelen, "Problems in cell kinetics and the early detection of disease," in Reliability and biometry, F. Proschan and R. J. Serfling (Eds.) Philadelphia, Society for Industrial and Applied Mathematics, 701-726 (1974).

[115] M. Zelen, "Theory of early detection of breast cancer in the general population," Technical Report 31, SUNY at Buffalo, 1975.

[116] M. Zelen and M. Feinleib, "On the theory of screening for chronic diseases," Biometrika 56, 601-614 (1969).

BIOGRAPHICAL SKETCH

Abdollazim Hoshyar was born on April 24, 1948, in Boroujerd, Iran, where he received his elementary and high school education. In June, 1965, he graduated from Bahre-olom High School with high honors. In September, 1966, he began undergraduate work in Pahlavi University. In June, 1971, he graduated first in his class with the degree Bachelor of Science in Mechanical Engineering.

Upon graduation, he was employed at the Arya-Mehr Mill plant in Isfahan, Iran, and after completion of two years of military service there he accepted the position of the Director of Vocational Centers of the National Iranian Society for Rehabilitation of the Disabled in Tehran. In March, 1975, Pahlavi University awarded him a scholarship for post-graduate work in the United States. He started graduate work in industrial engineering at the University of Florida. In December, 1976, he received the degree of Master of Science in Industrial and Systems Engineering.

He married Mehri Nakhjevan on August 24, 1971. His daughter Hani was born on November 27, 1973, in Tehran. Abdollazim Hoshyar is a member of Alpha Pi Mu and Tau Beta Pi honorary societies.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ralph W. Swain, Chairman
Associate Professor and Associate Director
Health Systems Research Division

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Thom J. Hodgson, Co-chairman
Associate Professor of
Industrial and Systems Engineering

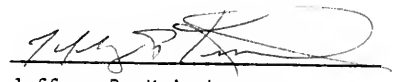
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Kerry E. Kilpatrick
Associate Professor and Director,
Health Systems Research Division

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gary J. Koehler
Associate Professor of Management

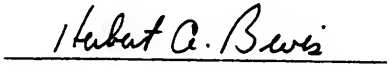
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Jeffrey P. Krischer
Associate Professor of
Industrial and Systems Engineering

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1978



Dean, College of Engineering

Dean, Graduate School

UNIVERSITY OF FLORIDA
3 1262 08666 282 1